

CARACTÉRISATION DE LA VARIABILITÉ ET DES DISTANCES GÉNÉTIQUES DES
BOVINS DE RACE CANADIENNE, SUISSE-BRUNE ET HOLSTEIN À L'AIDE DU
POLYMORPHISME DES CASÉINES ET DE MARQUEURS MICROSATELLITES

par

Hugo Yann Laliberté

mémoire présenté au Département de biologie en vue
de l'obtention du grade de maître en sciences (M.Sc.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, Janvier 1998



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*

Our file *Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-40599-0

Le 2 septembre 1998 , le jury suivant a accepté ce mémoire dans sa version finale.
date

Président-rapporteur: M. Richard Blouin
Département de biologie

Membre: M. Claude Déry
Département de biologie

Membre: M. Ryszard Brzezinski
Département de biologie

Membre: M. Denis Petitclerc
Agriculture Canada

SOMMAIRE

La race Canadienne est la plus ancienne race bovine de l'Amérique du Nord. Les individus de cette race se sont reproduits entre eux de façon isolée jusqu'à ce qu'on effectue des croisements avec la race Suisse Brune dans les années 70. Dans le but d'évaluer la diversité génétique des races Canadienne, Holstein et Suisse-Brune ainsi que les distances génétiques qui les séparent, le polymorphisme des caséines $\alpha S1$, β et κ ainsi que le polymorphisme de 14 marqueurs microsatellites (ILSTS005, ETH225, INRA35, ETH 152, INRAK, MM12, INRA63, INRA5, HEL1, INRA32, INRA72, HEL5, MM8, INRA16), uniformément distribués dans le génome, ont été caractérisés. Les individus des races Holstein (n=37), Canadienne (n=19) et Suisse-Brune (n=23) étaient non apparentés jusqu'à la deuxième génération. Le génotypage des caséines a été réalisé en grande partie à l'aide de gel en focalisation iso-électrique. De plus, suite à l'amplification par PCR des séquences microsatellites, des gels de séquençage utilisant une technique de détection par fluorescence ont été utilisés pour le génotypage.

Ainsi, le locus caséine $\alpha S1$ est monomorphe en race Canadienne et Holstein où on ne retrouve que le variant B. Une faible fréquence du variant C (0,06) est retrouvée en race Suisse-Brune. Le locus β comporte 3 allèles (A1, A2 et B) alors qu'on en détecte 2 (A et B) au niveau du locus κ . Concernant les loci microsatellites, un total de 95 allèles ont été détectés chez les 80 animaux analysés avec 6,8 allèles en moyenne par locus. Des taux d'hétérozygotie théorique de 0,66, 0,61 et 0,62 qui ne sont pas significativement différents entre eux ($P > 0,1$) ont été obtenus pour les races Canadienne, Suisse-Brune et Holstein, respectivement. L'échantillonnage de la race Canadienne présente un excès d'hétérozygotes par rapport à l'équilibre de Hardy-Weinberg, la race Suisse-Brune présente un déficit, alors que la race Holstein est en équilibre. Enfin, des phénogrammes consensus de 500 ré-échantillonnages, construits selon la méthode du Neighbor Joining et illustrant les distances génétiques standard de Nei et de Cavalli Sforza démontrent que la race Canadienne et Suisse Brune sont plus proches l'une de l'autre par rapport à la Holstein, avec des valeurs de bootstrap de 58% et 89%, pour chaque type de distance, respectivement.

REMERCIEMENTS

Je tiens à remercier tous ceux qui ont rendu possible, par leur soutien technique et moral, la réalisation de ce travail.

Denis Petitclerc et Claude Déry, les co-directeurs de cette recherche, ainsi que Richard Blouin et Ryszard Brzezinski, conseillers, qui par leurs conseils judicieux m'ont permis de bien m'orienter dans la réalisation de ce travail;

Jasmin Brochu, Danny Labrèque, Lynn Lessard et Gilles Robitaille pour leur aide et leur support moral;

Christiane Hansen pour l'échantillonnage effectué sur les trois races d'intérêt;

Patrice Martin et Christine Leroux (Doudou) dont la collaboration fut essentielle et pour leur accueil dans leur laboratoire. Marie-Françoise Mahé, Jean-Pierre Furet pour leur aide et tous les copains et copines du labo de Jouy-en-Josas. Et, bien sûr, je remercie (très très beaucoup) Katayoun Moazami-Goudarzi pour toute l'aide apportée tout au long de mes stages.

TABLE DES MATIERES

SOMMAIRE	II
REMERCIEMENTS.....	III
TABLE DES MATIERES	IV
LISTE DES TABLEAUX.....	VII
LISTE DES FIGURES	VIII
CHAPITRE 1 - INTRODUCTION	1
CHAPITRE 2 - REVUE DE LITTERATURE	3
2.1 Race Canadienne.....	3
2.1.1 Historique.....	3
2.1.2 Caractéristiques.....	3
2.2 Variabilité génétique	4
2.3 Marqueurs protéiques: les lactoprotéines.....	5
2.3.1 Description des lactoprotéines chez les bovins.....	5
2.3.2 Polymorphisme des lactoprotéines	5
2.4 Marqueurs génomiques	7
2.4.1 Caractéristiques des microsatellites	7
2.4.1.1 Caractéristiques génétiques.....	8
2.4.1.2 Caractéristiques techniques.....	9
2.5 Structure et variabilité génétique des populations	9
2.5.1 Fréquences alléliques	9
2.5.2 Variabilité et structure des populations.....	10
2.5.2.1 Indices de variabilité	10
2.5.2.2 Équilibre de Hardy-Weinberg	12
2.5.2.3 Structure des populations (Fstatistiques)	14
2.6 Variabilité inter-races.....	17
2.6.1 Distances génétiques	17
2.6.1.1 Distance de Cavalli-Sforza (1967)	18
2.6.1.2 Distance de Nei	19

2.6.2 Principe de construction de phénogrammes.....	20
2.6.3 Fiabilité des arbres	21
CHAPITRE 3 - MATÉRIELS ET MÉTHODES.....	22
3.1 Échantillonnage.....	22
3.1.1 Préparation des échantillons de lait.....	22
3.1.2 Extraction d'ADN.....	24
3.2 Analyse du polymorphisme des caséines.....	24
3.2.1 Électrophorèse à focalisation isoélectrique (IEF)	24
3.2.1.1 Principe de la méthode.....	24
3.2.1.2 Préparation du gel	25
3.2.1.3 Migration des échantillons.....	25
3.2.1.4 Coloration du gel.....	25
3.2.2 Autres méthodes.....	26
3.2.2.1 SSCP (single-strand conformation polymorphism)	26
3.2.2.1.1 Principe	26
3.2.2.1.2 Protocole	27
3.2.2.2 Méthode du ACRS (amplification created restriction sites)	27
3.2.2.2.1 Principe	27
3.2.2.2.2 Protocole	29
3.3 Analyse du polymorphisme des microsatellites.....	29
3.3.1 Principe	29
3.3.2 Choix des microsatellites	31
3.3.3 Amplification PCR	34
3.3.4 Gels d'électrophorèse	37
3.3.4.1 En Fluorescence	37
3.3.4.2 En Radioactivité.....	38
3.3.5 Analyse des résultats.....	38
3.3.5.1 En Fluorescence	38
3.3.5.2 En radioactivité	39

3.4 Méthodes statistiques	39
CHAPITRE 4 - RÉSULTATS	40
4.1 Polymorphisme des caséines.....	41
4.2 Polymorphisme des microsatellites.....	45
4.2.1 Fréquences alléliques	45
4.2.2 Taux d'hétérozygotie.....	45
4.2.3 Équilibre de Hardy-Weinberg.....	49
4.2.4 Analyse des F statistiques	51
4.2.5 Distances génétiques.....	53
CHAPITRE 5 - DISCUSSION	57
5.1 Polymorphisme des caséines.....	57
5.2 Polymorphisme des microsatellites.....	58
5.2.1 Variabilité et structure des populations.....	58
5.2.2 Distances génétiques.....	60
CHAPITRE 6 - CONCLUSION	63
BIBLIOGRAPHIE.....	64

LISTE DES TABLEAUX

1. Nombre d'échantillons collectés par race.....	23
2. Liste des amorces utilisées pour la méthode du SSCP	28
3. Liste des amorces utilisées pour la méthode du ACRS	30
4. Liste des microsatellites utilisés.....	35-36
5. Taux d'hétérozygotie observé et théorique non biaisé des loci caséiques.....	43
6. Comparaison des fréquences alléliques observées aux loci caséiques pour chaque race aux fréquences alléliques publiées dans la littérature.....	44
7. Nombre d'allèles obtenus par marqueur microsatellite et par race	46
8. Taux d'hétérozygotie observé et théorique non biaisé des loci microsatellites.....	48
9. Probabilités associées au test d'équilibre de Hardy-Weinberg.....	50
10. Structuration des races	52
11. F_{is} obtenus pour chaque race.....	52
Annexe 1. Fréquences alléliques des loci microsatellites	71
Annexe 2. Distances génétiques de Cavalli- Sforza.....	73
Annexe 3. Distances génétiques de Nei	73
Annexe 4. Identification des animaux utilisés	74

LISTE DES FIGURES

1. Schéma général du séquenceur 373A	32
2. Exemple de l'image d'un gel obtenu avec GeneScan	33
3. Histogramme des fréquences alléliques de la caséine α S1, β et κ	42
4. Exemple de distribution de fréquences alléliques de quelques microsatellites	47
5. Phénogramme consensus des distances Cavalli Sforza entre les races canadiennes, françaises et africaines selon la méthode du Neighbor-Joining	54
6. Phénogramme consensus des distances Nei entre les races canadiennes, françaises et africaines selon la méthode du Neighbor-Joining	55

CHAPITRE 1 - INTRODUCTION

L'agriculture moderne met fortement l'emphasis sur l'intensification des productions animales. Ainsi, les pratiques modernes se sont traduites par la spécialisation de quelques races bovines répondant mieux aux enjeux économiques actuels, au détriment de beaucoup d'autres races qui voient leurs effectifs diminuer. De plus, des techniques comme l'insémination artificielle qui sont certes très efficaces pour augmenter la productivité des troupeaux peuvent accélérer l'augmentation de la consanguinité. En plus de diminuer la variabilité génétique, la consanguinité multiplie les chances de voir l'expression de gènes récessifs létaux et a souvent des conséquences négatives sur la performance (Falconer, 1989; Hudson et Van Vleck, 1984). Quoi qu'il en soit, l'augmentation à courts termes de la productivité se fait souvent au détriment de la variabilité génétique. Ainsi, depuis la fin des années 70, de plus en plus de programmes de conservation ont été mis en place afin de mieux gérer la diversité génétique. En effet, étant donné les contraintes économiques actuelles, un équilibre entre les besoins de conservation des ressources et les besoins d'amélioration génétique doit être trouvé. L'idéal serait de préserver un nombre minimum de races tout en gardant un nombre maximum de variabilité. Ceci suppose des moyens plus précis d'investigation et d'évaluation de cette variabilité, permettant d'approfondir nos connaissances des relations entre races.

Le présent travail s'intéresse particulièrement à une race unique en Amérique du Nord, la race Canadienne. En effet, elle présente un intérêt certain étant donné son originalité et sa rareté. L'étude qui fait l'objet de ce présent mémoire s'intéresse à deux autres races: à la race Holstein, très largement répandue au Canada, et la race Suisse-Brune d'importance économique moindre. Étant donné l'état des races étudiées, deux hypothèses peuvent être formulées. D'abord, on peut supposer que la race Canadienne présente moins de variabilité génétique et une plus forte consanguinité que la race Holstein, étant donné son effectif limité. De plus, il est connu que des croisements ont été effectués entre la race Suisse-Brune et la race Canadienne. La deuxième hypothèse est donc que ces deux races sont génétiquement plus proches l'une de l'autre. Les objectifs du travail qui a été accompli ont donc été d'évaluer la variabilité génétique de ces trois races d'intérêt ainsi que les distances génétiques qui les séparent. Dans ce mémoire, nous exposerons donc dans une première partie les résultats obtenus après caractérisation de ces trois races à partir du polymorphisme des caséines et,

dans une deuxième partie, les résultats concernant le polymorphisme des marqueurs de type microsatellite. Finalement, nous présenterons les conclusions et perspectives de ce travail.

CHAPITRE 2 - REVUE DE LITTÉRATURE

2.1 Race Canadienne

2.1.1 Historique

La race Canadienne est la plus ancienne race bovine de l'Amérique du nord. Elle descend des anciennes races autochtones françaises provenant de la Bretagne et de la Normandie. De plus, elle semble être apparentée à la Jersey, la Guernsey et à la Kerry (Fortin, 1940). Les premières importations régulières de bovins en Nouvelle France s'effectuèrent de 1608 à 1610. Dès lors, à l'exception de quelques importations, ces bovins se reproduisirent entre eux de façon isolée pendant environ un siècle et demi (Bernier, 1995). Ainsi, dans les années 1800, malgré l'importation et le croisement de vaches Canadiennes avec des races britanniques, la presque totalité des troupeaux québécois étaient de race pure Canadienne (Couture, 1909). Cependant, elle fut rapidement marginalisée par l'expansion des races Holstein et Ayrshire et progressivement remplacée. En 1969, des croisements dans certains troupeaux de race Canadienne avec des animaux de race Suisse Brune américaine permirent de diminuer la consanguinité et d'améliorer la productivité (Bernier, 1995). Actuellement, la population de la race Canadienne compte environ 2000 têtes. La condition de la race est considérée comme critique selon les normes de l'American Livestock Breeds. Cet organisme considère qu'une race est en situation critique si le nombre d'entrées par année est inférieur à 200 et si la population totale est inférieure à 2000. Or, la race Canadienne a présenté seulement 209 enregistrements en 1995 (CLRC, 1995).

2.1.2 Caractéristiques

Les individus de la race Canadienne sont de petite taille. Le poids des femelles varie de 450 à 500 kg tandis que les mâles peuvent peser de 700 à 800 kg. La Canadienne est le plus souvent noire ou brune foncée (Bernier, 1995). Elle est bien adaptée au climat canadien et ne nécessite pas une nourriture particulière. Elle est donc facile à élever. La Canadienne produit un bonne quantité de lait compte tenu de sa petite taille et de sa consommation de nourriture (Bernier, 1995). Ainsi, la race Canadienne fait partie de ces races "rustiques" qui présentent souvent des adaptations intéressantes à des conditions particulières. En effet, les races

majoritaires sont en général adaptées à des conditions optimales. Ces races perdent souvent leurs avantages lorsqu'elles sont placées dans des conditions d'élevage moins favorables (Simon, 1984).

2.2 Variabilité génétique

Le terme variation se traduit intuitivement par le fait que tous les individus ne sont pas semblables entre eux. Cette différence se situe à différents niveaux et peut être liée à différents facteurs. La variation peut exister entre les individus mais également entre populations de différentes régions. On parle alors de variation de groupe ou de variation géographique. De plus, une part de cette variation n'est pas liée à l'hérédité mais est simplement due aux diverses réponses phénotypiques d'un même génotype à des conditions environnementales ou autres différentes. Bien sûr, il est important de distinguer la variabilité non-héréditaire de la variabilité héréditaire, c'est-à-dire la variabilité génétique. En effet, la variabilité génétique des populations se manifeste par l'existence, en de nombreux locus, de différentes formes de ces mêmes locus. Toutes les formes différentes d'un locus sont appelées des allèles. De plus, la variabilité est due à la multiplication de ces différents états alléliques par mutation, par recombinaison intragénique, etc. Ainsi, la variabilité génétique persiste dans les populations de génération en génération (Solignac et al, 1995).

La variabilité génétique peut être quantifiée en utilisant le concept de fréquence allélique (Hartl, 1988). Dans un groupe d'individus, la fréquence d'un allèle donné est simplement la proportion au locus considéré de tous les allèles de ce type. La notion de fréquence allélique est définie plus en détails dans la section 2.5.1. De nombreux marqueurs polymorphes, c'est-à-dire des marqueurs existant sous plusieurs états alléliques, peuvent être utilisés pour évaluer la variabilité génétique. Le polymorphisme morphologique (dans le cas des bovins: couleur des poils, forme et couleur des cornes, etc.) est un exemple de marqueur qui est souvent utilisé pour classer les races. Mais ce type de marqueurs est peu informatif et imprécis puisqu'un phénotype peut être associé à plusieurs gènes. De plus, certains de ces phénotypes peuvent s'exprimer de façon variable. Le polymorphisme des protéines et de l'ADN présente un plus grand intérêt. Ces types de polymorphisme seront traités plus en détails dans les sections suivantes.

2.3 Marqueurs protéiques: les lactoprotéines

Un certain nombre de marqueurs protéiques polymorphes peuvent être utilisés pour caractériser la variabilité génétique des races bovines et les distances génétiques qui les séparent (Baker et Manwell, 1980). Il s'agit le plus souvent de protéines sériques, érythrocytaires ou lactoprotéiques. Le polymorphisme protéique est caractérisé à l'aide de techniques d'électrophorèse. Les paragraphes suivants décrivent les différentes lactoprotéines retrouvées chez les bovins ainsi que leur polymorphisme.

2.3.1 Description des lactoprotéines chez les bovins

Généralement, on classe la portion protéique du lait en deux groupes: les protéines du lactosérum (soit l' α -lactalbumine et la β -lactoglobuline) et les caséines. Dans le lait des ruminants, on retrouve quatre caséines: l' α S₁, l' α S₂, la β et la κ . Les caséines sont les protéines majeures du lait. La concentration globale de celles-ci varie entre 25 et 32 g/l chez les bovins selon les races (Grosclaude, 1988). Les caséines sont des phosphoprotéines qui ont la propriété de former des micelles. D'ailleurs, c'est la coagulation ou la précipitation de ces micelles qui permet la fabrication des fromages. Chez les bovins, les gènes codant pour les différentes caséines se regroupent sur le même chromosome (Grosclaude et al, 1964, 1965, 1978; Treadgill et Womack, 1990). On remarque donc une très forte liaison entre ces quatre loci, contrairement aux gènes codant pour les protéines du lactosérum qui sont situées sur des chromosomes séparés (Grosclaude, 1988).

2.3.2 Polymorphisme des lactoprotéines

De nombreux travaux ont permis, à l'aide de différentes techniques d'électrophorèse, d'évaluer le polymorphisme des protéines du lait (Ng-Kwai-Hang et Grosclaude, 1992). Ainsi, la β -lactoglobuline présente environ 6 variants (A, B, C, D, F et G) parmi les populations de bovins (Aschaffenburg et Drewry, 1955; Bell, 1962, 1981b; Grosclaude et al., 1966), tandis que l' α -lactalbumine en compte 3 (A, B et C) (Bell et al., 1981a; Blumberg et Tombs, 1958). Pour ce qui est des caséines, on retrouve 4 variants (A, B, C, D) (Thompson et al., 1962; Grosclaude et al., 1966; Bell et al., 1981c) de la caséine α S₁ où B et C sont les plus fréquents de façon générale. Il est à noter que le variant A est retrouvé principalement en race Holstein (Kiddy et al., 1968; Li et Gaunt, 1972). Par contre, la caséine α S₂ est peu

polymorphe puisqu'on retrouve un allèle très largement majoritaire (le variant A) ainsi que deux allèles très rares, les variants C et D, ce dernier qu'on observe en particulier chez certaines races françaises (Grosclaude et al., 1976, 1979). La caséine β a 7 variants (A_1 , A_2 , A_3 , B, C, D et E) (Aschaffenburg et al., 1961, 1968; Peterson et Kopfler, 1966; Voglino, 1972) où les allèles A_1 et A_2 sont les plus répandus chez les bovins. Finalement, on observe 2 variants de la caséine κ (A et B) (Grosclaude et al., 1965) qui sont universellement répandus chez toutes les races bovines étudiées.

D'un point de vue moléculaire, la grande majorité des variants d'une même caséine diffèrent entre eux par la substitution d'un ou plusieurs acides aminés. Dans seulement deux cas, la différence est due à une délétion (la caséine αS_1 A et l' αS_2 D) (Ng-Kwai-Hang et Grosclaude, 1992). Ces différences dans la séquence de la chaîne d'acides aminés peuvent affecter les propriétés physico-chimiques des protéines comme, par exemple, le mécanisme de fixation des groupements phosphates dans la formation des micelles de caséines et ainsi affecter la fabrication des fromages (Grosclaude, 1988). Dans cet optique, il est raisonnable de penser que le polymorphisme des protéines du lait puisse influencer directement ou indirectement certains traits de production (quantité de lait produit, composition lipidique et protéique du lait, etc.). C'est pour cette raison que certains chercheurs se sont intéressés à utiliser les lactoprotéines comme "marqueurs" (Martin et Grosclaude, 1993). Dans cette approche, on tente de trouver une corrélation entre la présence de tel variant d'une lactoprotéine et un QTL (quantitative trait loci), c'est-à-dire un trait de production. Le but de cette approche est de pouvoir éventuellement sélectionner les porteurs des allèles qui semblent favorables puisqu'associés à un trait de production élevé. À l'heure actuelle, les résultats sont plutôt contradictoires et souvent non-comparables entre eux. Cependant, les chercheurs semblent d'accord sur le fait que les variants B de la caséine κ et A de la β -lactoglobuline soient associés à un contenu protéique plus élevé (Ng-Kwai-Hang et al., 1986, 1987; Ng-Kwai-Hang et Grosclaude, 1992; McLean et al., 1984; Rampilli et al., 1988).

Dans un autre ordre d'idées, le polymorphisme protéique donne certaines informations sur la variabilité génétique des races. Cependant, le polymorphisme protéique décrit de façon limitée le polymorphisme réel de l'ADN. En effet, l'étude des différents variants des lactoprotéines, par exemple, nous renseigne seulement sur les régions codantes de l'ADN (soit 4% du gène de la caséine κ , par exemple) qui, généralement, représente une faible proportion du gène. De plus, étant donné la dégénérescence du code génétique, le quart

environ des mutations dans les régions codantes n'ont pas d'effet sur la structure finale de la protéine. Finalement, les techniques d'électrophorèse utilisées pour caractériser le polymorphisme des protéines les séparent selon leur charge nette. Ainsi, une substitution d'acide aminé sur trois seulement se traduit par une variation de la charge nette. Alors, de façon globale, le polymorphisme protéique décrit seulement de 2 à 3% du polymorphisme réel de l'ADN (Mercier et Grosclaude, 1993). D'où l'intérêt d'utiliser, pour évaluer la diversité génétique, un marqueur qui décrit mieux le polymorphisme du génome dans son ensemble, soit les microsatellites.

2.4 Marqueurs génomiques

De façon générale, le polymorphisme de l'ADN peut être classé en trois types: les variations ponctuelles, les réarrangements de séquences et les variations du nombre de répétitions de séquences anonymes. Une proportion considérable (30 à 40%) du génome des mammifères est composée de séquences répétées. Ces séquences sont elles mêmes classées en trois catégories: les LINES (long interspersed elements), les SINES (short interspersed elements) (Singer, 1982) et les ADNs satellites. C'est dans ce dernier groupe que se retrouvent les microsatellites. Les ADNs satellites présente une organisation particulière. Il s'agit de motifs variants en taille d'un à plusieurs nucléotides et répétés un certains nombre de fois en tandem. En fait, quatre classes de séquences satellites sont distinguées empiriquement en fonction de la longueur de l'unité répétée: les satellites (plusieurs milliers de bases), les midisatellites (quelques centaines de bases), les minisatellites (quelques dizaines de bases) et enfin les microsatellites (quelques nucléotides).

2.4.1 Caractéristiques des microsatellites

Les microsatellites sont actuellement les marqueurs de choix dans plusieurs domaines. Mis à part en génétique des populations où ils permettent l'étude de la variabilité génétique, la consanguinité et le flux génétique des populations (Buchanan et al., 1994; MacHugh et al., 1994, 1997; Moazami-Goudarzi et al., 1997), ils sont très largement utilisés en tant que marqueurs en cartographie (Bishop et al., 1994). Les microsatellites sont également utilisés pour l'identification de gènes d'intérêt médical ou agronomique (QTL: quantitative trait loci).

2.4.1.1 Caractéristiques génétiques

La séquence nucléotidique qui constitue un microsatellite est composée de répétitions en tandem de trimères, dimères ou même monomères. C'est le nombre de répétitions du motif qui varie selon les allèles. Hamada et al. (1982) furent les premiers à signaler la présence de séquences (TG) répétées à des centaines de copies dans la levure et des milliers de fois chez les vertébrés. Les microsatellites sont très répandus puisqu'on les retrouve chez tous les eucaryotes étudiés jusqu'à maintenant (Tautz, 1989; Weber and May, 1989; Quellar et al., 1993). Ils sont généralement bien distribués dans le génome avec des variations selon les espèces. On retrouve, par exemple, un microsatellite (TG)_n à tous les 120 à 180 Kb (Steffen et al., 1993; Vaiman et al., 1994) chez les bovins tandis que la répartition de ce même type de microsatellite est de 1 à tous les 30 Kb chez l'homme (Stalling et al., 1991). Les microsatellites sont donc très nombreux à l'intérieur du génome (de 30000 à 100000 suivant les espèces) et de nature variable, c'est-à-dire que le type de microsatellite ((TG)_n, (CA)_n, etc.) que l'on retrouve le plus fréquemment varie d'une espèce à l'autre. Ainsi, les microsatellites de type (TG)_n sont les plus fréquents chez les bovins, ce qui n'est pas le cas chez l'homme, le rat ou le chien, par exemple. Enfin, les microsatellites sont des marqueurs très polymorphes par rapport aux autres marqueurs étudiés jusqu'à présent (polymorphismes protéiques, groupes sanguins, RFLP). En effet, 80% des séquences microsatellites isolées sont polymorphes. De plus, on estime que le nombre moyen d'allèles par microsatellite est de 6 (Vaiman et al., 1994) à 10 (Bishop et al., 1994).

Le rôle et l'origine des microsatellites ne sont pas, à ce jour, clairement établis. L'apparition de nouveaux allèles pourrait s'expliquer par le glissement de l'ADN polymérase durant la réplication ou par des crossing over inégaux (Schlötterer et Tautz, 1992). Deux modèles de mutation sont proposés: le SMM (stepwise mutation model) et le IAM (infinite allele model) (Rienzo et al., 1994; Shriver et al., 1993). Dans le cas du SMM, qui semble celui qui convient le plus aux microsatellites, les mutations se font par étapes où il y a un gain ou une perte d'une seule répétition. Le IAM, pour sa part, implique un nombre infini d'état allélique où des échanges inégaux amènent un très grand nombre d'allèles de longueurs différentes.

2.4.1.2 Caractéristiques techniques

Mis à part leurs propriétés génétiques, les microsatellites présentent des intérêts techniques considérables. En effet, le génotypage des microsatellites est basé sur l'utilisation de la technique PCR (polymerase chain reaction), procédure relativement simple et rapide. De plus, en combinant la PCR et la détection par fluorescence, plusieurs systèmes peuvent être étudiés simultanément. Enfin, la technique demande une quantité d'ADN très faible. Ces caractéristiques techniques favorisent ainsi la réalisation d'études de populations à grande échelle.

2.5 Structure et variabilité génétique des populations

L'analyse des données tirées d'une étude du polymorphisme de marqueurs protéiques ou de types microsatellites nécessite une approche statistique particulière. Le but de cette approche est de caractériser la variabilité et la proximité génétique des populations d'intérêt. D'une part, une première analyse statistique nous renseigne donc sur la structure et la variabilité génétique des populations. D'autre part, des calculs de "distances génétiques" permettent de quantifier les différences entre les populations. Cette section décrira donc les méthodes statistiques utilisées pour interpréter les résultats.

2.5.1 Fréquences alléliques

La principale valeur qui est à la base de l'analyse statistique de la variabilité génétique d'une population est la fréquence allélique, soit la proportion de tous les allèles donnés d'un même type à un certain locus dans cette population (Hartl, 1988). Les fréquences alléliques sont calculées par comptage direct du nombre d'animaux hétérozygotes puis du nombre d'animaux homozygotes pour l'allèle en question. Ainsi, la fréquence d'un allèle dans un échantillon est égale à deux fois le nombre de génotypes homozygotes pour cet allèle (car chaque homozygote porte deux copies de l'allèle) plus le nombre des génotypes hétérozygotes comportant cet allèle (car chaque hétérozygote porte une copie) divisé par deux fois le nombre total d'individus dans l'échantillon (parce que chaque individu porte deux allèles à ce locus). Formellement, la fréquence P_{ikx} de l'allèle i au locus k dans la population x est:

$$P_{ikx} = \frac{2(n_{ii}) + n_i}{2N}, \text{ où}$$

n_{ii} est le nombre d'animaux homozygotes pour l'allèle i au locus k ,
 n_i est le nombre d'animaux hétérozygotes pour l'allèle i au locus k
et N est le nombre total d'animaux typés au locus k avec

$$\sum_{i=1}^{l_k} P_{ikx} = 1, \text{ où}$$

l_k est le nombre d'allèles au locus k .

La suite de l'analyse de la variabilité génétique consiste à évaluer le taux d'hétérozygotie, l'équilibre d'Hardy-Weinberg et la structure des populations via l'analyse des F statistiques.

2.5.2 Variabilité et structure des populations

2.5.2.1 Indices de variabilité

La comparaison directe des fréquences alléliques n'est pas particulièrement facile à réaliser. Il existe donc des paramètres susceptibles de donner, par une valeur globale, les informations les plus importantes. On retrouve trois paramètres: le nombre moyen d'allèles par locus, le pourcentage de loci polymorphes et le taux moyen d'hétérozygotie (Solignac et al., 1995).

Le nombre moyen d'allèles par locus est calculé pour l'ensemble des locus étudiés pour chaque population. Cependant, le nombre absolu d'allèles trouvés dans un échantillon dépend de la taille de ce dernier: plus elle augmente, plus le nombre d'allèles pour le même nombre de locus sera élevé, des allèles rares étant révélés. Ainsi, ce paramètre présente un intérêt relativement limité. Le pourcentage de loci polymorphes varie également en fonction de la taille de l'échantillon. Les allèles rares ne seront détectés que si la taille de l'échantillon est suffisamment grande. Ainsi, si on classe de façon trop stricte les loci dans une catégorie soit monomorphe ou polymorphe, on risque de les classer dans l'une ou l'autre catégorie selon la taille de l'échantillon ou les hasards de l'échantillonnage. Pour éviter cet inconvénient, on se fixe un seuil de fréquence correspondant à la probabilité de détecter les allèles rares, soit 5%.

Un locus sera polymorphe si l'allèle second en fréquence a une fréquence supérieure à 5%, ou si l'allèle le plus fréquent a une fréquence inférieure à 95%.

Finalement, le taux d'hétérozygotie observée est, pour chaque locus, la fréquence des génotypes hétérozygotes observés dans l'échantillon. Un taux d'hétérozygotie théorique peut aussi être calculé à partir des fréquences alléliques déterminées pour chaque locus à l'aide de la formule suivante:

$$H = 1 - \sum P_i^2$$

où P_i est la fréquence du i^e allèle à ce locus. Cette hétérozygotie calculée est une valeur théorique qui correspond à l'hétérozygotie de la population, en la supposant à l'équilibre de Hardy-Weinberg (cette notion sera abordée dans la section suivante). Une moyenne est ensuite calculée pour chacun de ces taux, observé et théorique, pour l'ensemble des loci.

Le taux moyen d'hétérozygotie est l'indice le plus satisfaisant de la variabilité génétique. Sa valeur numérique dépend du nombre de loci polymorphes et de la structure génotypique de chacun d'eux. Pour un locus donné, la valeur de l'hétérozygotie est maximale lorsque la fréquence des allèles est identique et s'accroît avec le nombre d'allèles. Ce maximum est de $(n-1)/n$ pour n allèles. Cet indice de variabilité génétique peut être généralisé dans les cas où le taux d'hétérozygotie n'a pas de sens comme tel, comme dans le cas des organismes haploïdes par exemple. On parle alors de *diversité génique* qu'on définit comme étant la probabilité de tirer, au hasard, deux allèles différents à un même locus (Nei 1978). Ce paramètre est valable quel que soit le mode de reproduction et la degré de ploïdie. L'estimation non biaisée est:

$$H = \frac{2n(1 - \sum P_i^2)}{2n - 1}$$

où n est le nombre d'individus étudiés. Nei (1978) propose d'utiliser cette hétérozygotie non biaisée lorsque le nombre d'animaux testés est faible. De plus, on peut noter une corrélation positive entre le pourcentage de loci polymorphes et le taux d'hétérozygotie. En effet, plus la proportion de locus polymorphe est grande dans une population, plus les individus sont susceptibles d'être hétérozygotes. Le taux d'hétérozygotie dépend donc du type de marqueurs utilisés.

Les indices que nous avons définis peuvent donc décrire, jusqu'à un certain point, la variabilité génétique des individus ainsi que des populations que forment ces individus. Bien que la variabilité génétique d'un individu demeure inchangée tout au long de sa vie (exception faite d'éventuelles mutations), la constitution génétique d'une population, quant à elle, présente une possibilité de variation dans le temps. En effet, les populations subissent des forces qui déterminent leur constitution génétique et tendent à la maintenir ou à la modifier. Le modèle de Hardy-Weinberg, dont il est question dans la section suivante, nous permet de préciser les facteurs pouvant modifier les constitutions génétiques des populations.

2.5.2.2 Équilibre de Hardy-Weinberg

En 1908, un mathématicien anglais, G.H. Hardy et un médecin allemand, W. Weinberg ont formulé une loi connue sous le nom de la loi de Hardy-Weinberg. Selon cette loi, les fréquences alléliques restent stables de génération en génération dans une population diploïde idéale et ne dépendent que des fréquences de la génération initiale. Une telle population est dite en "équilibre de Hardy-Weinberg". De plus, les fréquences génotypiques ne dépendent que des fréquences alléliques. Dans ce modèle, une population diploïde idéale possède les propriétés suivantes:

1. Elle est de grande taille, idéalement de taille infinie. Elle n'est donc pas soumise à la dérive génétique;
2. La rencontre des gamètes se fait au hasard. Ceci implique que les gamètes s'unissent indépendamment de leur génotype (pangamie) et que l'union des individus des deux sexes se fait également au hasard (panmixie), c'est-à-dire qu'aucun choix de phénotype ou de génotype ne préside à la formation de couples. Cette propriété est souvent appelée: "hypothèse panmictique";
3. Il n'y a pas de migration. Aucune copie allélique n'est apportée de l'extérieur;
4. Il n'y a pas de mutations au sein de la population;
5. Elle ne subit pas de sélection;
6. Les générations ne sont pas chevauchantes, c'est-à-dire que les individus de chaque génération meurent avant que les membres de la génération suivante atteignent l'âge reproducteur.

Selon le modèle de Hardy-Weinberg, les fréquences des différents génotypes se calculent de la façon suivante. Considérons, par exemple, un locus à 2 allèles (A et a) possédant respectivement des fréquences p et q où $p + q = 1$. Pour qu'un individu soit homozygote AA, il faut qu'il ait reçu un allèle A de ses 2 parents. Si les gamètes s'unissent au hasard, cet événement se réalise avec la probabilité:

$$P(AA) = p \cdot p = p^2$$

Le même raisonnement s'applique pour le génotype aa et

$$P(aa) = q \cdot q = q^2$$

Enfin, pour le génotype hétérozygote Aa, deux cas sont possibles. L'individu peut avoir reçu l'allèle A de son père et l'allèle a de sa mère ou l'inverse. La probabilité d'obtenir un hétérozygote Aa est donc:

$$P(Aa) = p \cdot q + q \cdot p = 2pq$$

Ainsi, dans une population idéale, les proportions de Hardy-Weinberg sont données par

AA	Aa	aa
p^2	$2pq$	q^2

Cette situation est généralisable à un locus avec plusieurs allèles A_1, A_2, \dots, A_k . Les fréquences des homozygotes $A_i A_i$ sont égales à

$$f(A_i A_i) = p_i^2$$

et celles des hétérozygotes $A_i A_j$ sont égales à

$$f(A_i A_j) = 2p_i p_j$$

La notion d'équilibre de Hardy-Weinberg présente des conséquences théoriques importantes. Ainsi, il est possible de prédire, dans une population en équilibre de Hardy-Weinberg, les fréquences des différents génotypes à partir des seules fréquences alléliques. De plus, la ségrégation mendélienne aléatoire des chromosomes préserve la variabilité génétique des populations. Puisque les fréquences alléliques demeurent stables au cours du temps, une population diploïde idéale n'évolue pas. Seules les violations des propriétés de la population idéale permettent le processus évolutif, soit un changement des fréquences alléliques au sein

de la population. Donc, une perturbation de l'équilibre de Hardy-Weinberg peut indiquer l'effet d'un facteur évolutif tel que la consanguinité, la sélection, la migration, etc..

Pour vérifier qu'une population est en équilibre de Hardy-Weinberg à un locus donné, on teste, la plupart du temps à l'aide de la loi du χ^2 , la disparité des effectifs de génotypes observés aux effectifs de génotypes théoriques. Le logiciel utilisé pour effectuer ces calculs sera décrit dans le prochain chapitre.

Ainsi, la conséquence de l'action des facteurs évolutifs est de faire varier le taux d'hétérozygotes de la population par rapport à l'équilibre de Hardy-Weinberg. Il devient alors intéressant de pouvoir quantifier ces écarts entre le taux d'hétérozygotes observé et le taux à l'équilibre. De plus, deux populations d'une même espèce peuvent être soumises à des facteurs évolutifs différents et voir leur constitution génétique se différencier l'une de l'autre, ce qui aura également un effet sur le taux d'hétérozygotie et sur l'équilibre de Hardy-Weinberg. Tous ces écarts par rapport à l'équilibre de Hardy-Weinberg peuvent être quantifiés à l'aide de la statistique hiérarchique des coefficients F.

2.5.2.3 Structure des populations (Fstatistiques)

On distingue trois niveaux de complexité dans une population subdivisée: les individus (I), les sous-populations (S) et la population totale (T). Dans notre contexte, nos trois races représentent les sous-populations et l'ensemble de ces trois races représente la population totale. L'hétérozygotie de chacun de ces trois niveaux est définie par les paramètres suivants: H_I , H_S et H_T (Wright, 1978).

H_I est l'hétérozygotie moyenne des individus sur l'ensemble des sous-populations. H_I représente également l'hétérozygotie moyenne observée, pour l'ensemble des gènes (ou locus) d'un individu. C'est aussi la probabilité d'hétérozygotie en un locus pris au hasard. Ainsi, si H_i représente l'hétérozygotie observée dans la i^e sous-population, on aura, pour x sous-populations:

$$H_I = \sum_i^k H_i / x$$

H_S est l'hétérozygotie attendue par individu pour chaque sous-population (race), en la supposant à l'équilibre de Hardy-Weinberg. H_S représente également l'hétérozygotie attendue dans une sous-population supposée à l'équilibre de Hardy-Weinberg où P_i est la fréquence du i^e allèle. Soit pour la S^e population:

$$H_S = 1 - \sum_i^k P_{i,s}^2$$

On notera \bar{H}_S la moyenne des H_S sur les x sous-population:

$$\bar{H}_S = \sum_S^k H_S / x$$

Enfin, H_T représente l'hétérozygotie attendue par individu, en supposant la population globale à l'équilibre de Hardy-Weinberg. C'est aussi l'hétérozygotie attendue si toutes les sous-populations étaient regroupées en une seule unité panmictique. Si l'on note \bar{P}_i la fréquence moyenne de l'allèle A_i sur l'ensemble des x sous-populations, on obtient:

$$H_T = 1 - \sum_i^k \bar{P}_i^2$$

Trois indices sont définis à partir de ces hétérozygoties: F_{IS} , F_{ST} et F_{IT} . Ces trois indices mesurent l'écart de l'hétérozygotie par rapport à l'équilibre de Hardy-Weinberg à différents niveaux. Le premier indice, F_{IS} est défini par la relation:

$$F_{IS} = \frac{\bar{H}_S - H_I}{\bar{H}_S}$$

Cet indice mesure la réduction éventuelle d'hétérozygotie des individus à l'intérieur de leur sous-population. Il est également appelé *coefficient de consanguinité*. Dans des conditions où il y a consanguinité (lien de parenté entre individus possédant un ou plusieurs ancêtres communs), l'indice F_{IS} est positif et indique un déficit en hétérozygote. Évidemment, l'indice sera égale à zéro si les sous-populations sont à l'équilibre de Hardy-Weinberg. En revanche, si F_{IS} est négatif, les sous-populations présentent un excès d'hétérozygotes.

Entre les sous-populations et la population totale, l'effet de la subdivision est exprimé par un indice similaire:

$$F_{ST} = \frac{H_T - \bar{H}_S}{H_T}$$

F_{ST} est appelé indice de fixation et correspond à la réduction d'hétérozygotie dans les sous-populations, liée aux différences de fréquences alléliques moyennes. F_{ST} est nul lorsque toutes les sous-populations ont les mêmes fréquences alléliques et sont à l'équilibre. Dans le cas contraire, l'effet Wahlund implique que H_T soit plus grand que \bar{H}_S et donc F_{ST} sera positif. Ici, il convient d'expliquer que l'effet Wahlund (1928) est une réduction d'hétérozygotie de la population totale qui s'observe lorsqu'il existe une divergence des fréquences alléliques entre les sous-populations. Ces différences de fréquences alléliques moyennes sont généralement induites par l'effet de la dérive génétique. L'indice de fixation mesure donc la différenciation, l'effet de subdivision entre populations.

Enfin, la réduction d'hétérozygotie globale entre l'individu et la population globale théorique est donnée par l'indice:

$$F_{IT} = \frac{H_T - H_I}{H_T}$$

Les trois indices sont liés par la relation:

$$(1 - F_{IT}) = (1 - F_{IS})(1 - F_{ST})$$

Si toutes les sous-populations sont bien à l'équilibre de Hardy-Weinberg, on a $F_{IS} = 0$, et donc $F_{ST} = F_{IT}$. Si elles sont toutes à l'équilibre de Hardy-Weinberg et ont les même fréquences alléliques, F_{IS} et F_{ST} sont nuls et F_{IT} l'est aussi. La division en sous-populations n'existe plus en fait et la population globale est à l'équilibre de Hardy-Weinberg.

Comme il a été mentionné plus haut, l'indice F_{ST} permet de quantifier le degré de diversification génétique entre races. D'autres indices ont été proposés pour mesurer la distance génétique entre différents groupes. Ces notions de distance génétique seront développées dans la prochaine section.

2.6 Variabilité inter-races

Nous avons choisi deux fonctions de distance: celle de Cavalli-Sforza (1967) et celle de Nei (1974). Une fois les distances entre races calculées, il reste à les représenter graphiquement sous forme de phénogrammes. Deux méthodes de constructions de phénogrammes ont été retenues: la méthode UPGMA (Sokal et Michener, 1958) et la méthode du Neighbor-joining (Saitou et Nei, 1987). De plus, nous avons utilisé le test du bootstrap (Felsenstein, 1985) afin de vérifier la fiabilité des arbres obtenus. Les paragraphes suivants décriront ces méthodes.

2.6.1 Distances génétiques

Le but des calculs de distances génétiques est de quantifier les différences entre des groupes d'individus (des races dans notre cas). Chaque groupe est défini par un certain nombre de critères (les fréquences alléliques, par exemple) à partir desquels la proximité ou l'éloignement entre ces groupes est estimé. La notion de "distance" revient à réduire les données en une seule valeur qui caractérise la distance qui sépare les différents groupes.

La façon classique de calculer une distance génétique est de la représenter telle que nous la concevons dans un espace à trois dimensions. Ainsi, dans un espace tridimensionnel, un point possède trois coordonnées qui le positionne dans l'espace. Il est alors possible de calculer la distance euclidienne entre deux points en utilisant le théorème de Pythagore. Partant de ce principe, on peut se représenter une race, par exemple, comme un "objet mathématique" défini dans un espace à n dimensions dont les coordonnées sont les fréquences alléliques. Il s'agit alors d'extrapoler les principes de géométrie pour calculer un nombre qui prend le "sens" d'une distance. La notion de distance génétique comporte donc une part d'arbitraire et demeure quelque peu abstraite. De nombreux auteurs se sont intéressés aux propriétés que doit respecter une fonction d pour être considérée comme une fonction de distance. Gregorius (1984), par exemple, en énumère huit en tout. La plupart de ces propriétés s'appliquent intuitivement aux distances classiques telle que nous les concevons dans notre espace tridimensionnel. L'*inégalité du triangle* qui implique qu'il soit plus court de passer directement de X à Y que de passer par un intermédiaire Z , en est un exemple. D'autres propriétés présentent plutôt un intérêt pratique (que d soit bornée supérieurement) ou une appréciation théorique (que d soit maximum si les deux populations n'ont pas d'allèles en commun, que d ne dépende que des allèles communs aux deux populations, etc.). La

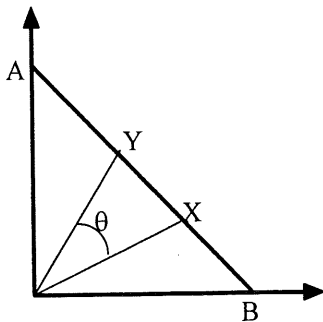
description détaillée de toutes les propriétés proposées par Gregorius dépasse l'objet de ce mémoire. Notons simplement que les nombreuses fonctions de distance qui existent dans la littérature ne respectent pas nécessairement toutes ces propriétés mais que celles-ci ne semblent pas toujours justifiables pour toutes études de populations. Il s'agit alors de tenir compte des propriétés particulières de chaque calcul de distance pour l'interprétation des résultats.

2.6.1.1 Distance de Cavalli-Sforza (1967)

Le principe générale, qui est à la base de la fonction de distance de Cavalli-Sforza, est le suivant. Prenons, par exemple, deux populations X et Y caractérisées par les vecteurs x et y avec:

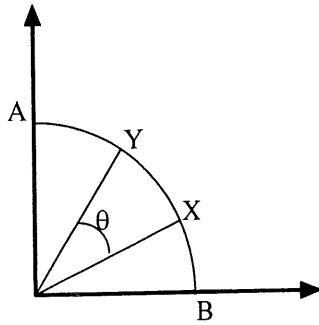
$$\sum_{i=1}^l p_{ix} = 1 \text{ et } \sum_{i=1}^l p_{iy} = 1$$

Ces vecteurs correspondent, dans un espace à l dimensions, à des points se déplaçant sur la portion d'hyperplan (si $l = 2$). Si $l = 2$, c'est-à-dire si on considère 2 allèles (A et B), par exemple, il s'agit d'un segment de droite AB situé dans le premier quadrant.



On peut songer à mesurer l'écart entre deux vecteurs (xy) par l'angle θ qu'ils forment. Mais un même écart entre x et y sur la droite AB correspond à un angle θ d'autant plus petit que x et y sont proches de A ou de B. En fait, pour que cette distance reste inchangée, il faut que l'angle θ reste le même. Ceci est vrai si les vecteurs (xy) sont de mêmes longueurs. Pour cela, il faut que ces points se déplacent sur une portion d'hypersphère. En effectuant un changement de variable qui remplace les éléments de x et y par leurs racines carrées, de nouveaux vecteurs sont obtenus, Ils correspondent à des points se déplaçant dans un espace à l dimensions sur la portion d'hypersphère de rayon 1.

Si $l = 2$, il s'agit de la portion AB du cercle de rayon 1 situé dans le premier quadrant.



Partant de ce concept, la distance entre X et Y (appelée d_{corde}) est calculée à l'aide des principes de la trigonométrie. La méthode de calcul est évidemment extrapolée dans le cas d'un espace à l dimensions où l représente le nombre de fréquences alléliques caractérisant les races étudiées. Les détails de ces calculs dépassent le cadre de ce mémoire. Enfin, Cavalli Sforza et Bodmer (1971) ont cherché à corréler cette distance à un paramètre de génétique des populations: la variance de Wahlund. Cette variance (f_θ) représente une mesure de l'hétérogénéité entre des populations évoluant indépendamment. Ils obtiennent alors la distance F_θ qui est la distance de Cavalli Sforza (1967):

$$F_\theta \approx \frac{4 \sum_{k=1}^r \left(1 - \sum_{i=1}^{l_k} \sqrt{P_{iX} P_{iY}} \right)}{\sum_{k=1}^r (l - 1)}.$$

2.6.1.2 Distance de Nei

L'utilisation des mesures de distances et de diversité de Nei se justifie par leur signification biologique (Nei, 1972, 1978; Nei et Roychoudhury, 1978). En effet, ces distances sont basées sur la probabilité d'identité des gènes et donc étroitement associées à la définition du coefficient de parenté. La probabilité de tirer dans deux populations X et Y le même allèle est:

$$I = \sum_{i=1}^{l_k} P_{iX} P_{iY}$$

Il s'agit de l'indice de similarité de Sneath (Sneath et Sokal, 1973). Nei (1972) propose de normaliser cet indice pour qu'il soit compris entre 0 et 1.

$$I = \frac{j_{XY}}{\sqrt{j_X j_Y}}$$

avec $j_{XY} = \sum_{i=1}^{l_k} P_{iX} P_{iY}$, $j_X = \sum_{i=1}^{l_k} P_{iX}^2$ et $j_Y = \sum_{i=1}^{l_k} P_{iY}^2$ où :

j_X est la probabilité de tirer dans la population X deux allèles identiques, j_Y est la probabilité de tirer dans la population Y deux allèles identiques et j_{XY} la probabilité d'identité d'un allèle tiré dans X et d'un allèle tiré dans Y. Ainsi, $I = 0$ quand les populations n'ont pas d'allèles en commun et $I = 1$ quand les populations sont identiques. Nei exprime ensuite la *distance standard* D par :

$$D = -\ln(I)$$

Ainsi, si $I=0$ alors $D=\text{infini}$ et si $I=1$ alors $D=0$. De plus, D correspond au nombre moyen de substitutions alléliques survenues à ce locus depuis la séparation des deux populations.

2.6.2 Principe de construction de phénogrammes

Les phénogrammes sont des diagrammes qui traduisent le degré de similitude entre les groupes. Ils reflètent une ressemblance globale, quelque soit son origine. En effet, le degré de similitude n'est pas nécessairement le reflet du degré de parenté. La ressemblance entre les groupes peut être due aussi bien au partage de caractères hérités d'un ancêtre commun proche qu'au partage de caractères hérités d'un ancêtre plus lointain (homologie). Ou bien, l'identité de caractères peut être due à l'apparition indépendante de caractères similaires (homoplasie). En fait, le phénogramme peut être considéré comme représentant correctement la *phylogénie* du groupe seulement si les taux d'évolution sont constants dans les diverses lignées et au cours du temps. Quoi qu'il en soit, on considère que plus la ressemblance entre deux unités évolutives, par exemple la race, est importante, plus la parenté entre elles a des chances d'être proche.

Il existe différentes méthodes phénétiques qui se distinguent par leurs hypothèses évolutives et leurs algorithmes. Nous avons utilisé les méthodes du Neighbor-Joining (NJ) (Saitou et Nei, 1987) et du Unweighted Pair Group Method Analysis (UPGMA) (Sokal et Michener,

1958). Ces méthodes sont dites agglomératives parce qu'elles regroupent les populations en partant des plus ressemblantes pour aller jusqu'aux moins ressemblantes. Ainsi, dans une première étape, les populations les plus ressemblantes sont recherchées et regroupées. Une nouvelle matrice de distances est alors calculée. Le mode de calcul de cette nouvelle matrice varie d'une méthode à l'autre. Dans la méthode UPGMA, chaque nouvelle distance est obtenue en calculant une moyenne pondérée par le nombre de populations considérées. Le processus est poursuivi jusqu'à ce que toutes les populations soient regroupées en une seule population hypothétique. La méthode du NJ est très semblable à la méthode UPGMA. De la même façon, dans une première étape, les deux unités évolutives (UE) les plus proches sont recherchées et regroupées. Considérons qu'il s'agit de X et Y. Une nouvelles matrice de distance est ensuite calculée, après avoir retiré X et Y pour les remplacer par l'unité évolutive hypothétique a (UEHa). Les distances de cette UEHa aux autres UE sont celles préconisées par Fitch et Margoliash (1967):

$$d_{aK} = \frac{1}{2}(d_{XK} + d_{YK} - d_{XY})$$

Le processus est poursuivi ainsi jusqu'à ce que toutes les UE soient agglomérées.

2.6.3 Fiabilité des arbres

Le test du Bootstrap est utilisé pour évaluer la confiance que l'on peut accorder à un arbre (Felsenstein, 1985; Wu, 1986)). Cette méthode suppose que les caractères ont évolué indépendamment les uns des autres et suivent une même loi de distribution. Il s'effectue par ré-échantillonnage des caractères avec remise. Un tirage au hasard avec remise de k caractères parmi les K constituant les données est effectué. Un arbre est construit après chaque ré-échantillonnage, ce processus étant réitéré 500 fois, par exemple. Un arbre de consensus synthétise les 500 arbres ainsi obtenus, permettant de visualiser leurs concordances. Dans la représentation donnée, chaque noeud (regroupement) est accompagné de son pourcentage d'apparition. En principe, un noeud est considéré comme significatif s'il apparaît dans 95% des arbres. Cependant, dans la littérature concernant l'étude de races et non d'espèces (Machugh et al., 1994; Bowcock et al., 1994), le seuil de stabilité du noeud est en général fixé à 75%, les valeurs inférieurs à 50% étant à considérer avec prudence.

CHAPITRE 3 - MATÉRIELS ET MÉTHODES

3.1 Échantillonnage

Des échantillons de sang et de lait de différents troupeaux de race Canadienne, Holstein et Suisse-Brune ont été prélevés (Tableau 1). Afin d'obtenir une représentation la plus fidèle possible de la variabilité génétique des races étudiées, les animaux ont été choisis de façon à ce qu'ils soient non-apparentés jusqu'à la deuxième génération (c'est-à-dire qu'ils n'aient ni parents, ni grand-parents en commun). Le choix des individus de la race Canadienne et Suisse-Brune a été réalisé par l'équipe de Christian Hansen (1997) en se basant sur les généalogies fournies par la Canadian Livestock Records Corporation (CLRC). Il est intéressant de signaler que les échantillons récoltés au sein de la race Canadienne représentent une très grande proportion des familles existantes. Enfin, le choix des échantillons appartenant à la race Holstein a été fait à l'aide des généalogies fournies par l'Association Holstein du Canada.

3.1.1 Préparation des échantillons de lait

La méthode qui a été utilisée la plupart du temps consiste à diluer les échantillons de lait dans une solution d'urée 8M (dilution 1/4, concentration finale de l'urée = 6N). Cependant, il a été nécessaire de lyophiliser une partie des échantillons pour faciliter leur transport. Le lait ainsi traité a été redilué dans de l'urée 8M. De plus, il est préférable dans certains cas de ne déposer qu'uniquement les caséines préalablement précipitées, pour ainsi faciliter la lecture des gels. En effet, les protéines du lactosérum (α lactalbumine et β lactoglobuline) migrent dans la même région que la β caséine (Seibert et al., 1985), ce qui complique la lecture. Le principe de cette méthode consiste à abaisser le pH du lait pour faire précipiter les caséines. Un volume de 1,5 ml de lait est pipeté dans un tube eppendorf et centrifugé à 12000 RPM pendant 2 min. Une grande proportion de gras est retirée puis 1 ml de lait est transféré dans un autre

Tableau 1. Nombre d'échantillons collectés par races

Canadienne	Suisse-Brune	Holstein
19	23	37

tube. Le lait est repris dans 50 µl d'acide acétique 20 % et incubé à 37 °C pendant 5 min. L'incubation terminée, 50 µl d'acétate de sodium 20 % sont ajoutés avant une deuxième incubation de 5 min à 37 °C. Les tubes sont ensuite centrifugés à 12000 RPM pendant 2 min. La couche de gras est de nouveau aspirée ainsi que le surnageant qui contient le lactosérum. Le culot de caséines est resuspendu dans 1 ml d'eau. On procède ensuite à une nouvelle centrifugation à 12000 RPM pendant 2 min. Le surnageant est de nouveau éliminé. Cette dernière opération est reprise en ajoutant 1 ml d'eau puis en centrifugeant. Une fois le surnageant retiré, on procède à la resolubilisation du culot de caséine. Après avoir ajouté 40 µl de NaOH 0,2 N au culot, celui-ci est resuspendu. Le volume de la solution est ensuite augmenté de 260 µl d'eau et de 100 µl de NaOH 0,2 N. Le tube est finalement vortexé puis le volume de la solution est complété à 1 ml.

3.1.2 Extraction d'ADN

L'extraction de l'ADN génomique a été fait à partir d'échantillons de sang. La technique du "Wizard Genomic DNA Purification Kit" de Promega a été utilisés selon le protocole proposé. La concentration ainsi que l'état de chacun des ADN extraits ont ensuite été vérifiés sur gel d'agarose 1% sur lequel a migré conjointement le marqueur de taille λ HindIII.

3.2 Analyse du polymorphisme des caséines

3.2.1 Électrophorèse à focalisation isoélectrique (IEF)

3.2.1.1 Principe de la méthode

Des gels d'électrophorèse à focalisation isoélectrique (IEF) ont été utilisés pour identifier les différents variants caséiques des individus échantillonnés. Ces variants diffèrent entre eux par leur point isoélectrique (pI), c'est-à-dire le pH où la protéine a une charge neutre. Ainsi, pour distinguer les variants, il s'agit de faire migrer les caséines dans un gradient de pH formé à l'aide d'ampholites sur un gel de polyacrylamide. Ainsi, la migration des protéines cesse lorsque celles-ci atteignent le pH correspondant à leur point isoélectrique.

Un appareil Multiphor II a été utilisé pour réaliser les gels d'électrophorèse IEF. Le protocole suivant s'inspire fortement de celui proposé par Seibert et al. (1985).

3.2.1.2 Préparation du gel

La migration des échantillons se fait sur un gel de polyacrylamide 5,4%. Lors de la préparation d'un gel, l'une des deux plaques de verre (celle qui sera en contact avec le gel) est traitée avec du diméthylchlorosilane en suivant les indications du fabricant (Pharmacia). Trois ampholites sont ajoutés à 17 ml d'une solution d'acrylamide, bis-acrylamide et d'urée (5,2%-0,16%-8M), soit 200 µl de l'ampholite de pH variant entre 4,2 et 4,9, 150 µl de l'ampholite pH 2,5-5 et 50 µl de l'ampholite pH 5-6,5. On ajoute ensuite 20 µl de TEMED et 40 µl d'une solution de persulphate d'ammonium 10% (préparation extemporanée) pour permettre la polymérisation du gel d'acrylamide. La solution est immédiatement injectée entre les deux plaques soit avec une seringue ou avec un cylindre gradué selon le type de plaques utilisées.

3.2.1.3 Migration des échantillons

Les échantillons dilués 1:4 dans l'urée 8M sont préparés à être déposés sur gel en ajoutant du mercaptoéthanol (concentration finale de 3% v/v). Le gel est libéré des plaques de verre et placé avec la feuille de plastique qui le supporte sur la plaque de migration de l'appareil Multiphor II. On place ensuite sur le gel les mèches préalablement trempées soit dans une solution basique (NaOH, 0,1 N) pour la cathode, soit dans une solution acide (H_3PO_4 , 0,1 N) pour l'anode. Une préfocalisation est effectuée afin de faire migrer les ampholites pour ainsi former le gradient de pH dans le gel. La puissance est réglée à 11 watts. On laisse migrer les ampholytes environ 5 min ou jusqu'à ce que le voltage atteigne 700 volts. Un volume de 6 µl de chacun des échantillons est alors déposé via de petits papiers filtre sur le gel. Une courte migration de 5 min à 11 watts permet de faire entrer les échantillons dans le gel. La migration s'effectue à 22 watts pendant 1 h 20 min, tout en maintenant la température à 15° C.

3.2.1.4 Coloration du gel

Une fois la migration terminée, le gel est placé dans une solution d'acide trichloroacétique (TCA) 10% pendant au moins 1 heure pour précipiter les protéines. Après avoir retiré le

TCA, le gel est recouvert de colorant (Bleu de comassie 0,16%, 30% MeOH, 10% acide acétique) et laissé sous agitation pendant 5 min environ. Le gel est ensuite décoloré pendant 2 heures par lavage dans une solution: 26% MeOH, 8% glycérol et 8% acide acétique. Le gel est ensuite séché ou placé dans une enveloppe de plastique.

3.2.2 Autres méthodes

Nous avons analysé le polymorphisme des caséines de certains échantillons dont le lait était manquant (tarissement de la vache choisie, mammite, etc.) à partir de leur ADN. Ainsi, après PCR (Polymerase Chain Reaction), ces échantillons ont été étudiés en utilisant les techniques de SSCP (single-strand conformation polymorphism) ou d'ACRS (amplification created restriction sites).

Le but de la PCR est d'obtenir rapidement un très grand nombre de copies d'un segment spécifique d'ADN. Cette technique est rendue possible grâce à l'emploi d'une ADN polymérase thermorésistante, la *Taq* polymérase. En effet, cette enzyme dont la température optimale d'action se situe aux environs de 70 °C, peut résister aux températures de dénaturation de l'ADN. De plus, la mise en oeuvre de la technique implique l'emploi de deux amorces oligonucléotidiques dont les séquences sont respectivement complémentaires à des séquences situées en amont et en aval du fragment que l'on désire amplifier. Ainsi, la technique PCR comporte trois étapes: la dénaturation de l'ADN, où la température du mélange réactionnel est portée à 94°C, l'hybridation des amorces (entre 50 et 65°C) et l'extension des brins d'ADN (72°C). Ces trois étapes forment un cycle qui est répété une trentaine de fois, ce qui permet la synthèse de plusieurs fragments doubles brins dans lesquels sont incluses, aux extrémités, les amorces d'extension.

3.2.2.1 SSCP (single-strand conformation polymorphism)

3.2.2.1.1 Principe

La technique du SSCP est basée sur le fait que la migration d'un fragment d'ADN simple brin sur un gel non-dénaturant est influencée par la séquence de ce fragment. Ainsi, deux fragments de taille identique migreront, en principe, plus ou moins vite sur un gel non dénatrant selon la conformation particulière que prendra ces fragments en fonction de leurs

séquences respectives. Ce principe est appliqué suite à l'amplification par PCR de la région contenant la mutation connue des variants à discriminer. Dans notre cas, les variants A et B du locus de la caséine K ont été différenciés grâce à cette technique.

3.2.2.1.2 Protocole

La première étape consiste à amplifier par PCR la région du génome contenant la mutation. Le volume réactionnel contient 1 µg d'ADN, 0,2 mM de dNTP (dATP, dCTP, dTTP et dGTP), 50 mM de MgCl₂, 1 nMol de chacune des amorces (voir tableau 2), 0,625 U de *Taq* Polymérase ainsi que du tampon PCR (Tris-HCl 10 mM, KCl 50 mM) pour un volume total de 25 µl. Le programme PCR est le suivant: l'ADN est d'abord dénaturée à 94°C pendant 3 min avant de subir 40 cycles d'amplification. Chaque cycle comprend une dénaturation de 30s à 94 °C suivi d'une hybridation à 55 °C pendant 30 s puis d'une élongation à 72 °C pendant 1 min. L'ADN est laissée encore 5 min à 72 °C après les cycles d'amplification.

La seconde étape consiste en la migration des échantillons dans un gel de polyacrylamide non dénaturant. Un µl du produit d'amplification est mélangé avec 20 µl de tampon de dépôt (LIS: 40% sucrose, 0,01% de bleu de bromophénol, 0,01% xylène cyanol). Après dénaturation à 97 °C pendant 3 min, les échantillons sont déposés sur un gel de polyacrylamide 12% (acrylamide-bisacrylamide 100:1) contenant du tampon TBE (Tris, Borate, EDTA). Un gel de retardement de 2 cm contenant 50% formamide est inclus pour prévenir complètement la réhybridation de l'ADN simple brin durant la phase initiale de l'électrophorèse. La migration est effectuée à 300 V pendant 16 h en maintenant la température du gel à 15 °C sur un appareil d'électrophorèse BioRad Protean II. Une fois l'électrophorèse terminée, le gel est coloré à l'argent.

3.2.2.2 Méthode du ACRS (amplification created restriction sites)

3.2.2.2.1 Principe

La méthode ACRS est elle aussi basée sur la technique PCR. Elle consiste à introduire artificiellement un site de restriction discriminant pendant l'amplification par PCR (Haliassos et al., 1989). La méthode peut s'appliquer à n'importe quelle mutation ponctuelle, à condition bien sûr qu'un site de restriction puisse être introduit par mutagenèse dirigée et que ce site

Tableau 2. Liste des amorces utilisées pour la méthode du SSCP

Nom de l'amorce	Locus	Position	Séquence 5'→3'
CASKS6	κcas	nt 10551	TGTGCTGAGTAGGTATCCTAGTTATGG
CASKA7	κcas	nt 11004	GCGTTGTCTTCTTTGATGTCTCCTTAG

soit spécifique à un des deux allèles que l'on veut discriminer. L'introduction du site de restriction dépend de l'amorce 3' qui contient un nucléotide non apparié. La distinction des variants est réalisée suite à la digestion du produit PCR et à la migration des fragments sur gel d'agarose. Ainsi les variants B et C de la caséine α S1 et les variants A1, A2 et B de la caséine β ont été être différenciés de cette façon (Lien et al., 1992, 1993).

3.2.2.2 Protocole

Deux réactions PCR sont réalisées. La première permet d'amplifier la région d'intérêt du génome. La seconde permet d'introduire le site de restriction. Le volume réactionnel de la première PCR contient 400 ng d'ADN, 0,2 mM de dNTP (dATP, dCTP, dTTP et dGTP), 1,5 mM de $MgCl_2$, 1 nMol de chacune des amorces, 0,625 U de *Taq* Polymérase ainsi que du tampon PCR 1X (Tris-HCl 10 mM, KCl 50 mM) pour un volume total de 25 μ l. La seconde PCR est identique à l'exception qu'elle est réalisée avec 6 μ l de la première PCR. La liste des amorces utilisées est présentée dans le tableau 3. Le programme est le même quelque soient les amorces. L'ADN est d'abord dénaturée à 94°C pendant 3 min avant de subir 40 cycles d'amplification. Chaque cycle consiste à une dénaturation de 30 s à 94 °C suivi d'une hybridation à 55 °C pendant 30 s puis d'une élongation à 72 °C pendant 1 min. L'ADN est laissée encore 1 min à 72 °C après les cycle d'amplification. On procède à la digestion des différents amplicons à l'aide de leurs enzymes respectives tel que décrit par le fabricant (New England Biolabs). Le tableau 3 présente les différentes enzymes de restriction à utiliser selon les amorces. Finalement, on procède à la migration des fragments sur gel d'agarose 3 % avec un voltage de 90 V pendant 40 min. Les bandes sont révélées aux U.V. grâce au bromure d'éthidium.

3.3 Analyse du polymorphisme des microsatellites

3.3.1 Principe

La variabilité génétique des individus se mesure à l'aide du nombre de répétitions du motif des microsatellites qui diffère d'un allèle à l'autre. Ainsi, pour repérer les allèles d'un microsatellite, il s'agit d'amplifier par PCR une région du génome qui le contient. La taille des amplicons est déterminée par électrophorèse en gel d'acrylamide. Deux approches, l'une fluorescente (nécessitant un automate d'analyse d'ADN) et l'autre radioactive, ont permis de

Tableau 3. Liste des amorces utilisées pour la méthode du ACRS

Paires d'amorces	Locus	Séquence 5'→3' *	Taille	Site de restriction introduit
CASAS1 (1)	αcas	ACAATTATACCAGCTGGATGCCTATC	344	
CASAS1 (2)		CACGCTCCACAGTTCCTGAGTAA		
CASAS1 (192)+	αcas	CCGGATCCTATTGGCTCTGAGAACGGTG	250	<i>HphI</i> dans le variant B
CASAS1 (2)		voir plus haut		
CASB (1)	βcas	CAGACACAGTCTCTAGTCTATCCCTT	390	
CASB (2)		TCTCTGGGATAGGGCACTGCTTTCT		
CASB(67)+	βcas	CCTGCAGAATTCTAGTCTATCCCTTCCCTGGGCCC	251	<i>TaqI</i> dans les variants A1 et B
		ATCG		<i>BstNI</i> dans les variants A1 et A2
CASB(122)		GAGTCGACTGCAGATTTTCAACATCAGTGAGAGT		
		CAGGCCCTG		

* La base qui est soulignée permet de créer le site ACRS.

mettre en évidence le polymorphisme des microsatellites. Ainsi, l'appareillage utilisé pour l'approche fluorescente est un automate d'analyse d'ADN de type 373 A (Applied Biosystem) composé d'un module d'électrophorèse et d'un logiciel le "GeneScan" développé sur "Macintosh". Il permet le dépôt de 36 échantillons différents par gel et la mise en évidence simultanée du polymorphisme de plusieurs microsatellites (une dizaine) par animal et par gel. Au niveau de l'automate, pendant la migration, les fluorophores sont excités par un rayon laser qui balaie la partie inférieure du gel (voir Figure 1). Ils émettent alors une lumière de longueur d'onde spécifique. Celle-ci passe à travers une roue à filtre avant d'être reprise par un tube photomultiplicateur qui la convertit en signal électrique. Les données brutes sont ensuite transmises à un ordinateur. Un logiciel, le GeneScan 373, reconstruit l'image du gel (voir Figure 2). Avec l'approche radioactive, un ensemble classique de matériel d'électrophorèse dans lequel 60 échantillons différents peuvent être déposés par microsatellite est utilisé. Le polymorphisme d'un seul (trois au maximum) microsatellite par animal et par gel peut être détecté. Par ailleurs, la détermination de la longueur exacte des différents allèles s'effectue en fluorescence, à partir de l'ADN du phage lambda digéré par l'enzyme *Pst*I, marqué par un fluorophore (rouge dans notre cas). Ce standard comigre avec les échantillons. En radioactivité, c'est la séquence du bactériophage M13, déposé dans des puits adjacents aux échantillons qui est utilisée.

Ces microsatellites ont été dénichés au niveau du génome bovin par différents chercheurs à l'aide d'une technique d'hybridation. En effet, des banques partielles d'ADN génomique ont été criblées avec des sondes oligonucléotidiques de synthèse complémentaires à la séquence des microsatellites recherchés. Les sondes qui ont été choisies par les différentes équipes sont, en général, des poly TG et TC. Les clones plasmidiques ayant intégré un insert contenant un microsatellite ont ensuite été séquencés. Finalement, des couples d'amorces oligonucléotidiques ont été définis à partir des séquences flanquant les microsatellites.

3.3.2 Choix des microsatellites

Le choix des microsatellites a été effectué en fonction de leurs caractéristiques génétiques (polymorphisme, localisation, étude préalable dans d'autres populations) et de leurs caractéristiques techniques (aptitude à l'amplification, lecture aisée des résultats). De plus, 11 des 14 microsatellites choisis font partie du panel de 30 microsatellites recommandé par la

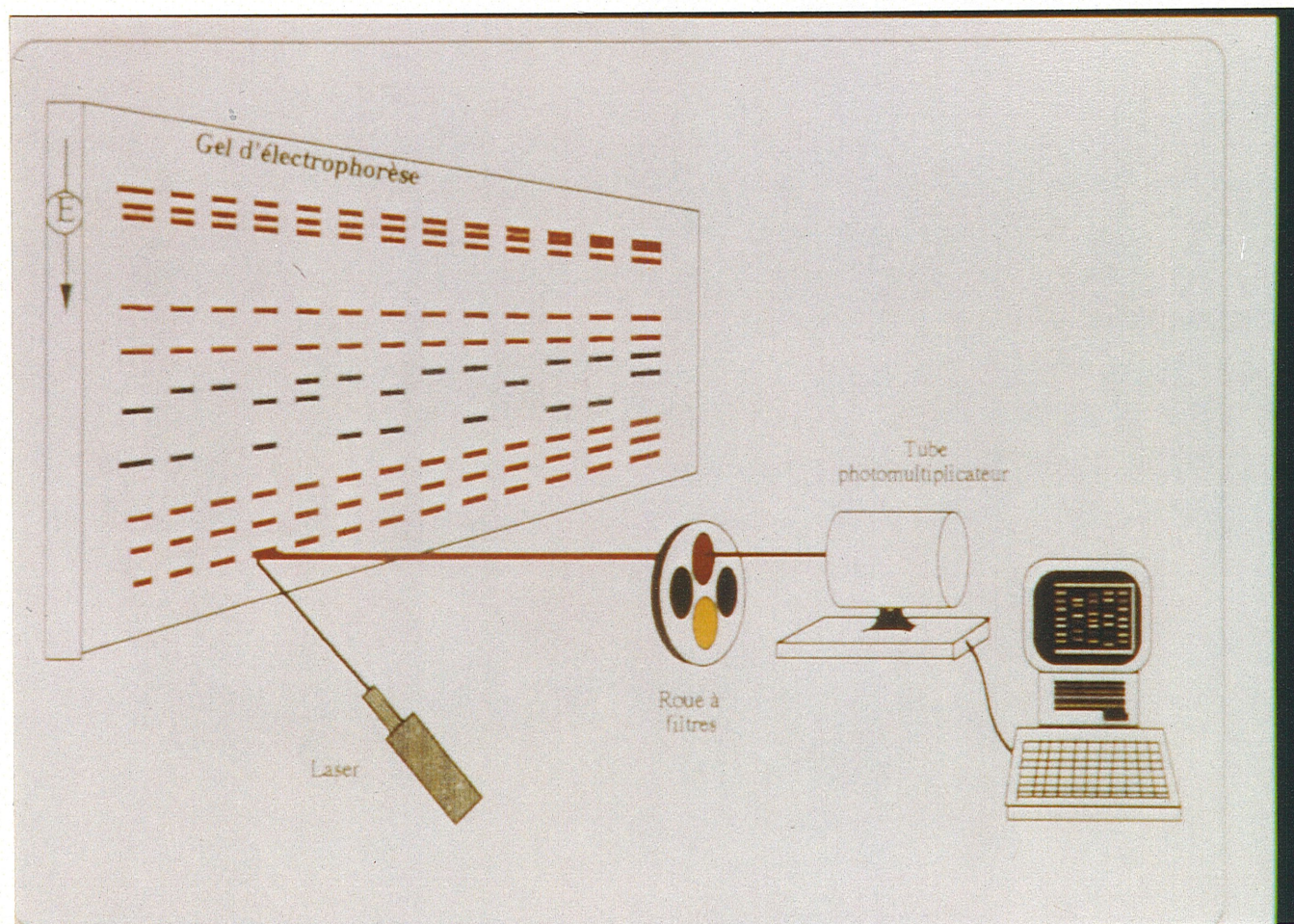


Figure 1. Schéma général du séquenceur 373A

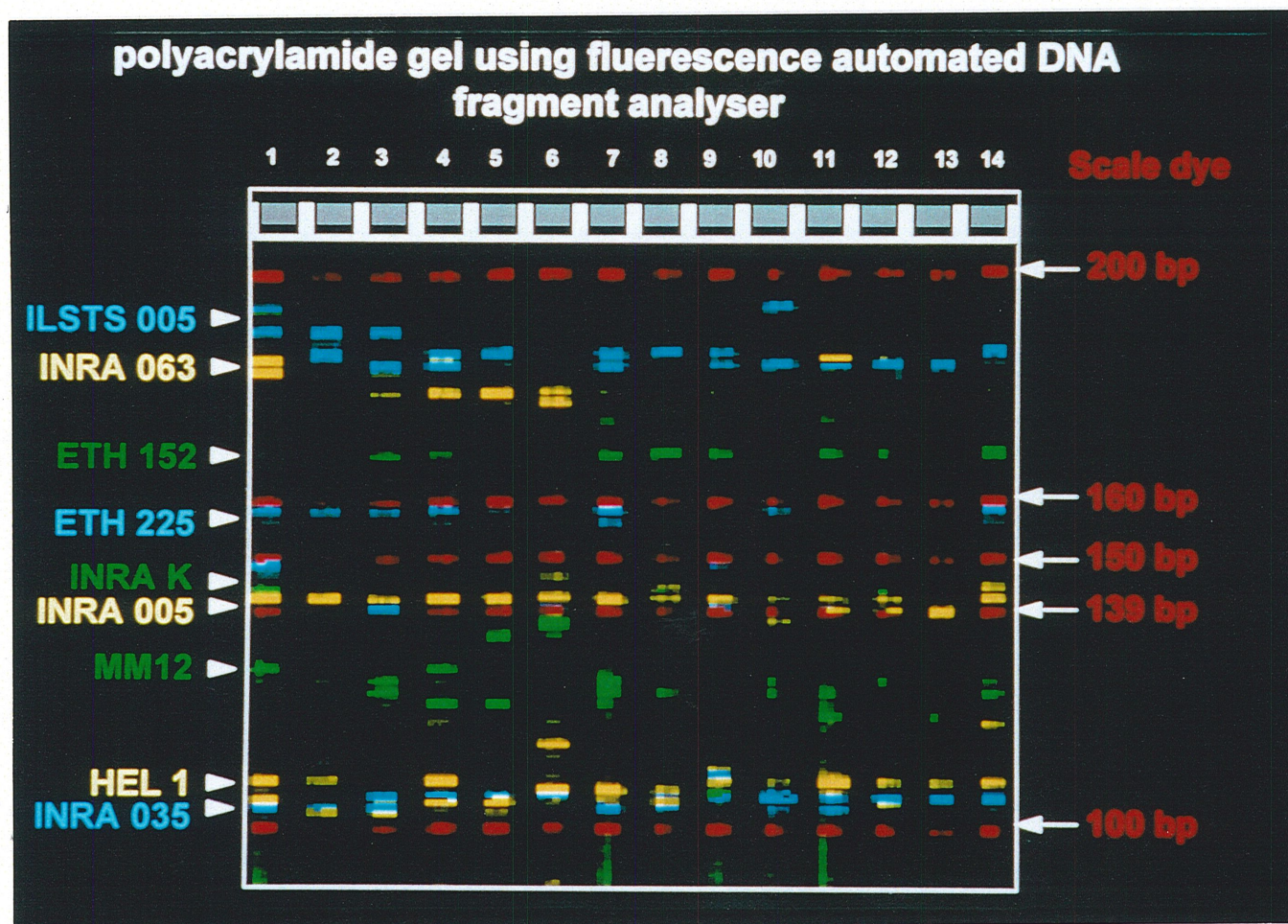


Figure 2. Exemple de l'image d'un gel obtenu avec GeneScan

Food and Agriculture Organisation (F.A.O.) (1993) ([HTTP://11dae.fao.org/dad-is/data/molecula/madad.htm](http://11dae.fao.org/dad-is/data/molecula/madad.htm)) pour l'analyse de la diversité des races bovines. Ces marqueurs correspondent également aux critères déterminés lors de la conférence de Dublin en 1995. Ils sont localisés sur des chromosomes différents afin de représenter l'ensemble du génome (voir Tableau 4). L'unité de répétition de la plupart des microsatellites que nous avons étudiés était de 2 bases. La mise au point des conditions de PCR des 14 couples d'amorces utilisés pour ce travail a été effectuée par K. Moazami Goudarzi (INRA, Jouy-en-Josas, France).

3.3.3 Amplification PCR

L'appareil PCR qui a été utilisé est de marque Perkin Elmer, modèle Cetus 9600. Un séquenceur automatique modèle 373A a également été utilisé.

Les produits d'amplification ont été marqués selon deux principes différents, soit en fluorescence ou soit en radioactivité (^{35}S). En fluorescence, seule une des deux amorces est marquée en 5'. Ceci implique que seulement un des deux brin d'ADN résultant de la PCR sera marqué et donc détecté au niveau du gel d'électrophorèse d'acrylamide. Quatre fluorophores de couleurs différentes peuvent être utilisés. Le volume réactionnel de la PCR contient 100 à 200 ng d'ADN, 125 μM de dNTP (dATP, dCTP, dTTP et dGTP), 1,5 mM de MgCl_2 , 15 pMol de chacune des amorces, 0,375 U de *Taq* Polymérase ainsi que du tampon PCR 1X (Tris-HCl 10 mM, KCl 50 mM) pour un volume total de 15 μl . La "PCR radioactive" quant à elle permet de marquer les deux brins d'ADN. En effet, au cours de cette PCR, un des désoxyribonucléotides (dATP) est substitué par son équivalent radioactif ($(\alpha\text{-}^{35}\text{S})\text{-ATP}$). L'épitope radioactif est donc incorporé tout au long de la synthèse d'ADN. Notons que seul le système d'amplification concernant le microsatellite INRA 16 a été utilisé en radioactivité. Le volume réactionnel de la PCR contient 100 à 200 ng d'ADN, 25 μM des trois désoxynucléotides: dCTP, dGTP et dTTP, 2,5 μM dATP et 1 μCi de $(\alpha\text{-}^{35}\text{S})\text{-dATP}$, 1,5 mM de MgCl_2 , 15 pMol de chacune des amorces, 0,375 U de *Taq* Polymérase ainsi que du tampon PCR 1X (Tris-HCl 10 mM, KCl 50 mM) pour un volume total de 15 μl .

Les conditions d'amplification sont identiques en fluorescence ou en radioactivité. Ainsi, l'ADN est dans un premier temps dénaturé 5 min. à 94 °C pour ensuite subir trente cycles d'amplification. Les tubes sont alors portés à 72 °C pendant 1 min. pour une dernière

Tableau 4. Liste des microsatellites utilisés

Nom	Amorces 5'→3'	Taille	Nombre d'allèles théoriques	Chromosome	Fluorophore	Références
MM8*	CCCAAGGACAGAAAAGACT CTCAAGATAAGACCACACC	148-122	14	Chr 2	4-tet (vert)	Mommens et al. (1994)
INRA72	CTTAACTCATTACCTCAACTG AGTGATTGAGCACATTGCGCAT	143-111	17	Chr 4	6-fam (bleu)	Vaiman et al. (1994)
ETH152*	TACTCGTAGGGCAGGCTGCCTG GAGACCTCAGGGTTCGTGATCAG	212-194	10	Chr 5	4-tet (vert)	Steffen et al. (1993)
INRAK	ATGCACCTTAACCTAATCC ACAATACACAAGCATACTAC	178-162	9	Chr 6	4-tet (vert)	Levesiel et al. (1994)
ETH225*	GATCACCTTGCCACTATTTCCT ACATGACAGCCAGCTGCTACT	165-137	15	Chr 9	6-fam (bleu)	Steffen et al. (1993)
MM12*	CAAGACAGGTGTTTCAATCT ATCGACTCTGGGGATGATGT	146-104	22	Chr 9	4-tet (vert)	Mommens et al. (1994)
ILSTS005*	GGAAGCAATGAAATCTATAGCC TGTTCTGTGAGTTTGTAAGC	193-175	9	Chr 10	6-fam (bleu)	Brezensky et al. (1993)

* Fait parti du pannel de Dublin, 1995

Tableau 4. Liste des microsatellites utilisés (suite)

Nom	Amorces 5'→3'	Taille	Nombre d'allèles théoriques	Chromosome	Fluorophore	Références
INRA32*	AAACTGTATTCTCTAATAGCTAC GCAAGACATATCTCCATTCTTT	192-162	16	Chr 11	6-fam (bleu)	Vaiman et al. (1994)
INRA005*	CAATCTGCATGAAGTATAAATAT CTTCAGGCATACCCTACACC	163-131	17	Chr 12	5-hex (jaune)	Vaiman et al. (1992)
HEL1*	CAACAGCTATTTAACAAGGA AGGCTACAGTCCATGGGATT	118-102	9	Chr 15	5-hex (jaune)	Kaukinen et Varvio (1993)
INRA35*	ATCCTTTGCAGCCTCCACATTG TTGTGCTTTATGACACTATCCG	127-101	14	Chr 16	6-fam (bleu)	Vaiman et al. (1994)
INRA63*	ATTTGCACAAGCTAAATCTAACC AAACCACAGAAATGCTTGGAAG	187-171	9	Chr 18	5-hex (jaune)	Vaiman et al. (1994)
HEL5*	GCAGGATCACTTGTTAGGGA AGACGTTAGTGTACATTAAC	195-175	11	Chr 21	4-tet (vert)	Kaukinen et Varvio (1993)
INRA16	ACGCAGACCTTAGCATAGGA GTCGCAATGAGTTGGACACAAC	155-135	11	Chr 27	marquage radioactif	Vaiman et al. (1994)

* Fait parti du pannel de Dublin, 1995

élongation pour finalement être refroidis à 4 °C. Chaque cycle comprend: 30 s de dénaturation à 94 °C, 30 s d'hybridation des amorces à 55 °C et 30 s d'élongation à 72 °C.

3.3.4 Gels d'électrophorèse

3.3.4.1 En Fluorescence

Avec l'automate d'analyse de l'ADN, il est possible de faire migrer dans le même puits des produits de PCR de tailles identiques mais marqués par des couleurs différentes et des produits de PCR marqués par la même couleur mais de tailles différentes. Il s'agit alors de réaliser la meilleure combinaison couleur taille des microsatellites. Ainsi, pour déterminer le nombre de marqueurs utilisables par piste, il faut tenir compte du nombre d'allèles et de la taille de chaque microsatellite.

Après PCR, quelques amplicons correspondant à chaque système sont déposés sur gel d'agarose 4% (2% agarose + agarose Nusive 2%). Ce premier gel permet de vérifier l'amplification et de quantifier sommairement la concentration du produit de la PCR. La présence des différentes bandes est révélée à l'aide de bromure d'éthidium en solution dans le gel d'agarose. Par la suite, des dilutions des différents produits de la PCR sont effectuées si nécessaire. En effet, pour faciliter la lecture et l'interprétation des résultats sur le séquenceur automatique, il est préférable de déposer des échantillons ayant une concentration à peu près identique pour chacun des systèmes.

Après dilution, pour chaque puits du gel d'acrylamide, un "pool" correspondant aux produits d'amplifications des différents microsatellites est constitué (mélange de 5 µl de chaque système). Concrètement, les systèmes d'amplification des microsatellites ILSTS005, ETH225, INRA035, ETH152, INRAκ, MM12, INRA063, INRA005 et HEL1 forment un premier groupe et sont déposés sur un même gel et les systèmes des microsatellites INRA032, INRA072, HEL5 et MM8 forment un deuxième groupe. Le tableau 4 indique la taille des amplicons de chaque système ainsi que la couleur du fluorophore des amorces.

Enfin, 5 µl de chaque "pool" est déposé dans d'autres cupules pour être séché et dénaturé par chauffage à 90°C pendant environ 14 min. Un volume de 5 µl d'un tampon de migration est ajouté à chaque cupule. Le tampon de migration est composé de 125 µl de formamide

déionisée, 31 μ l de dextran et 15,5 μ l de marqueur de taille (genescan 2500 rox, rouge). Les échantillons sont finalement chauffés de nouveau à 90°C pendant 3 min. et immédiatement gardés sur glace avant d'être déposés sur gel. La migration est effectuée à 40 W, 21 mA et 1500 V (voltage constant) pendant 8 à 9 heures.

3.3.4.2 En Radioactivité

Après 4 minutes de dénaturation à 90 °C, 7,5 μ l de la réaction de PCR sont déposés sur un gel vertical de polyacrylamide 5% dénaturant (urée 7 M). La migration est effectuée à 1500-1600 volts constant, pendant 1,5 à 4 heures suivant la taille de l'amplicon. Le gel est ensuite transféré sur papier Wathman 3MM, séché 1 heure à 80°C et autoradiographié. La durée d'exposition du film peut varier d'un jour à une semaine suivant l'intensité du signal.

3.3.5 Analyse des résultats

3.3.5.1 En Fluorescence

Avant de commencer la migration, des paramètres de migration, des paramètres de traitements, la feuille de route (nom des échantillons et systèmes analysés) et les paramètres d'analyses sont fixés. Pendant la migration, la collection et l'analyse des données nécessitent l'emploi de deux logiciels. Le premier logiciel (GS collection 672) construit un fichier qui contient les données brutes que l'automate d'analyse a enregistré pendant la migration des échantillons dans le gel. À la fin de la migration, le second logiciel (GS analysis 672) est automatiquement activé. Ce logiciel récupère le fichier de données collectées et construit l'image du gel selon les paramètres de traitement et la feuille de route. Le logiciel calcule la taille d'un fragment en effectuant une moyenne de deux régressions linéaires (une première régression linéaire sur les trois points suivants: la bande du marqueur lambda de taille inférieure au fragment et des deux bandes de tailles immédiatement supérieures; une deuxième régression linéaire sur les trois points suivants: une bande du marqueur lambda de taille supérieure au fragment et des deux bandes de tailles immédiatement inférieures). Une fois le gel analysé, il reste à faire correctement la correspondance entre la taille de l'amplicon et l'allèle dont il est question. Étant donné l'incertitude sur la mesure de la taille des différents fragments, le logiciel n'assigne pas toujours exactement, d'un gel à l'autre la même taille à un

allèle donné. La valeur arrondie de cette taille peut être tantôt paire tantôt impaire. Pour palier à ce problème, des échantillons dits témoins (déjà typés par des études préalables) sont déposés dans tous les gels de façon à permettre une standardisation de tous les résultats.

3.3.5.2 En radioactivité

Les résultats sont analysés manuellement à partir du film d'autoradiographie. La taille peut être déterminée à l'aide d'un marqueur. Des témoins dont on connaît déjà le génotype sont également utilisés pour déterminer rapidement la composition allélique des échantillons analysés.

3.4 Méthodes statistiques

L'approche statistique qui a été utilisée dans cette étude est composée de deux parties. La première partie nous renseigne sur la structure et la variabilité génétique des trois populations tandis que la seconde partie permet de quantifier et de représenter les relations qui existent entre ces trois races.

Les fréquences alléliques ont été calculées par comptage direct. Une première analyse, qui consiste à comparer la distribution des fréquences alléliques par locus et par race a été réalisée avec un test du χ^2 à l'aide du programme SAS (SAS Institute, 1989). Les taux moyens d'hétérozygotes (observés et théoriques non-biaisés) aux loci microsatellites de chaque race ont été comparés entre eux à l'aide d'une analyse de variance (SAS Institute, 1989). Enfin, les logiciels Genepop, Fstat, et Phylip ont été utilisés pour effectuer la plupart des analyses statistiques de nos données.

L'équilibre de Hardy-Weinberg a été vérifié à l'aide du logiciel Genepop (Raymond et Rousset, 1995). En effet, ce logiciel effectue un "test exact de Fisher" tel que décrit par Louis et Dempster (1987). De plus, le logiciel Genepop permet de tester l'indépendance entre les populations et leurs compositions alléliques. Ce test est réalisé selon la méthode développée par Raymond et Rousset (1994).

L'analyse des statistiques de F a été réalisée avec le logiciel Fstat (1984). Ce logiciel calcule les valeurs de F_{IS} pour chaque sous-population, de F_{ST} pour chaque locus ainsi que la valeur

globale pour tous les locus et enfin F_{IT} . De plus, ce logiciel teste les valeurs obtenues par Jacknife et calcule une déviation standard. Le Jacknife est une méthode de ré-échantillonnage. Il consiste, lorsqu'il est effectué au niveau de toutes les sous-populations, à retirer au hasard une sous-population pour ainsi former un nouvel échantillonnage sur lequel les calculs sont effectués. Cette opération est réalisée un certain nombre de fois. Le logiciel Fstat effectue également le Jacknife en tenant compte de tous les loci en retirant, cette fois, un locus au hasard. La méthode du Bootstrap est également utilisée pour ré-échantillonner les données ce qui permet de calculer un intervalle de confiance pour les valeurs de F_{IS} , F_{ST} et F_{IT} . Finalement, Fstat peut réaliser un test en effectuant des permutations d'allèles et ainsi calculer la probabilité que F_{IS} , F_{ST} et F_{IT} soient égales à zéro.

Enfin, Phylip (Felsenstein, 1993) est un ensemble de programmes qui permettent de réaliser, entre autre, des calculs de distances et des constructions de phénogrammes. Dans un premier temps, un ré-échantillonnage est effectué par la méthode du Bootstrap (500 fois) avec le programme *Seqboot*. Les distances de Nei et de Cavalli Sforza sont ensuite calculées par le programme *Gendist*. Puis, les phénogrammes sont construits avec le programme *Neighbor* selon les méthodes du Neighbor-Joining (Saitou et NEI, 1987) et du UPGMA (Sokal et Michener, 1958). Finalement, un arbre consensus, regroupant les 500 phénogrammes effectués selon l'une ou l'autre des méthodes précédentes, est construit avec le programme *Consense*. On obtient donc en tout quatre arbres consensus résultant des combinaisons des deux méthodes de calculs de distances et des deux méthodes de construction de phénogrammes.

CHAPITRE 4 - RÉSULTATS

4.1 Polymorphisme des caséines

Seuls les loci des caséines $\alpha S1$, β et κ ont été considérés puisque la caséine $\alpha S2$ est totalement monomorphe chez les trois races étudiées. En tout, 5 allèles ont été détectés. Le locus $\alpha S1$ est monomorphe pour les races Canadienne et Holstein puisqu'on ne retrouve que l'allèle B; la race Suisse-Brune ne présentant qu'une faible fréquence du variant C (0,06) à ce locus. Le locus β s'avère être le plus polymorphe avec 3 allèles (A_1 , A_2 et B) détectés chez les races Holstein et Suisse-Brune, la Canadienne ne présentant que les variants A_1 et A_2 . Enfin, on retrouve 2 allèles (A et B) au locus κ détectés chez les trois races. La répartition des fréquences alléliques des différents variants est illustrée dans la figure 3. Les taux d'hétérozygoties observés sont relativement faibles (de 0 au locus $\alpha S1$ à 0,59 au locus β chez la Holstein) étant donné le petit nombre d'allèles détectés par locus. Les valeurs d'hétérozygoties observées chez les trois races étudiées sont présentées dans le tableau 6. Les taux d'hétérozygoties observés ne sont pas significativement différents entre les trois races ($P > 0,1$), contrairement aux taux d'hétérozygoties théoriques ($P < 0,05$). Notons également que les valeurs des fréquences alléliques que nous avons obtenues sont dans l'ensemble en accord avec les valeurs tirées de la littérature (Ng-Kwai-Hang et Grosclaude, 1992; Swaisgood 1992) (voir Tableau 5). La race Suisse-Brune présente cependant une exception avec des valeurs légèrement différentes des valeurs de la littérature aux locus des caséines β et κ . Finalement, le test exact de Fisher indique que les trois races sont en équilibre de Hardy-Weinberg à chacun des locus caséiques. De plus, les trois races sont significativement différentes entre elles si on prend compte l'ensemble des loci (test d'indépendance entre les populations). Le locus de la caséine κ est celui qui différencie le mieux les races entre elles ($P < 0,001$ pour le locus κ , $P < 0,05$ au locus β et $P < 0,5$ au locus $\alpha S1$).

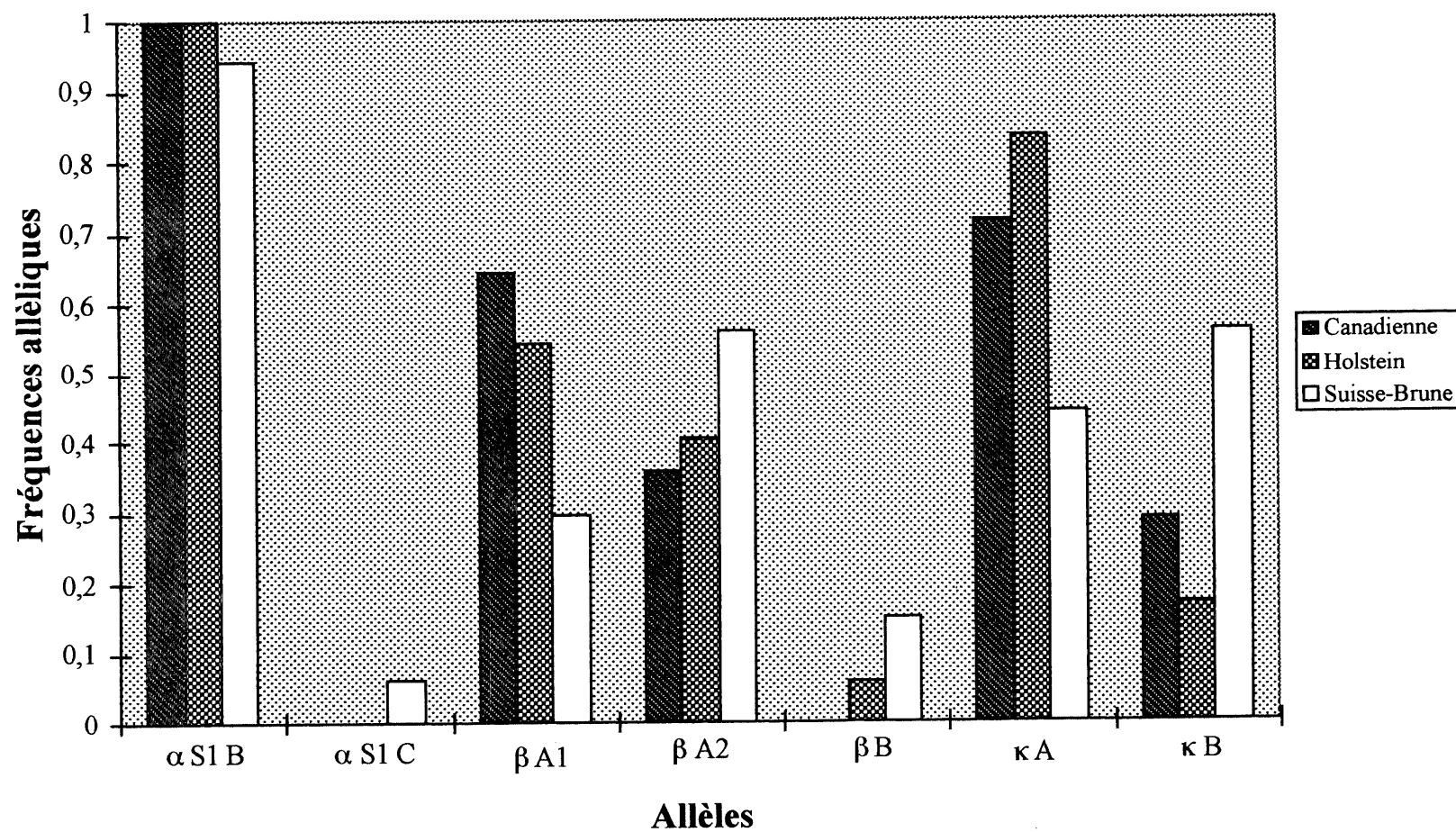


Figure 3. Histogramme des fréquences alléliques des caséines α S1, β et κ

Tableau 5. Taux d'hétérozygotie observé et théorique non-biaisé des loci caséiques

Locus	Races							
	Canadienne		Suisse-Brune		Holstein		Moyenne par locus	
	Observé	Théorique	Observé	Théorique	Observé	Théorique	Observé	Théorique
α SIcas	0,00	0,00	0,12	0,12	0,00	0,00	0,04	0,04
β cas	0,43	0,48	0,53	0,73	0,59	0,54	0,52	0,58
κ cas	0,37	0,42	0,55	0,49	0,27	0,27	0,40	0,4
Moyenne par race	0,27	0,30	0,40	0,45	0,29	0,27		

Tableau 6. Comparaison des fréquences alléliques observées aux locus caséiques pour chaque race aux fréquences alléliques publiées dans la littérature

Variant caséique	Races					
	Canadienne		Suisse-Brune		Holstein	
	Obtenues	Littérature*	Obtenues	Littérature**	Obtenues	Littérature**
α S1 B	1,000	na	0,941	0,980	1,000	0,940
α S1 C	0,000	na	0,059	0,020	0,000	0,060
β A1	0,643	0,585	0,294	0,150	0,5441	0,490
β A2	0,357	0,336	0,559	0,720	0,4118	0,490
β B	0,000	0,000	0,147	0,180	0,0441	0,010
κ A	0,714	0,790	0,441	0,700	0,8382	0,750
κ B	0,286	0,210	0,559	0,300	0,1618	0,250

* Ng-Kwai-Hang et Grocclaude, 1992

** Swaisgood, 1992

4.2 Polymorphisme des microsatellites

4.2.1 Fréquences alléliques

Les 14 loci microsatellites qui ont été étudiés se révèlent être tous polymorphes dans les trois races. Un total de 95 allèles ont été détectés chez les 80 animaux analysés. En considérant les trois races, on retrouve de 4 (INRA κ et INRA 072) à 11 (MM 12) allèles par locus (voir Tableau 7). Le nombre moyen d'allèles pour ces mêmes loci varie de 3 à 7. Enfin, le nombre moyen d'allèles par race varie de 4,8 ($\pm 1,5$) pour la race Suisse-Brune à 5,1 ($\pm 1,7$) pour les races Holstein et Canadienne.

Il est également intéressant de remarquer que la distribution des fréquences alléliques diffère d'un locus à l'autre pour une même race et d'une race à l'autre pour un même locus. La figure 4 montre des exemples de distribution de fréquences alléliques de quatre loci microsatellites. On y remarque une distribution des fréquences alléliques tantôt unimodale (INRA 032), tantôt bimodale (MM8) ou parfois même plus complexe. D'ailleurs, la distribution de fréquences alléliques de chaque locus est significativement différente d'une race à l'autre selon un test de X^2 ($P = 0,001$, à l'exception des microsatellites INRA 063, INRA 005 et MM 8 où $P = 0,05$) (résultats non présentés).

4.2.2 Taux d'hétérozygotie

Les valeurs de taux d'hétérozygotie observée et théorique non biaisée par locus et par race sont reportées dans le tableau 8. Étant donné le faible nombre d'animaux qui constitue l'échantillonnage de chacune de ces trois races, il est préférable de considérer l'hétérozygotie théorique non biaisée (également appelée diversité génétique non biaisée). En effet, puisque l'hétérozygotie théorique varie selon le nombre d'allèles détectés, elle peut être sous évaluée si l'échantillonnage est trop faible. Ainsi, la valeur moyenne de l'hétérozygotie théorique non

Tableau 7. Nombre d'allèles obtenus par marqueur microsatellite et par race

Locus	Races			Total par locus	Moyenne par locus
	Canadienne	Suisse-Brune	Holstein		
ILSTS005	4	3	3	6	3,33
ETH152	6	5	8	8	6,33
INRA35	5	4	3	5	4,00
ETH152	4	5	5	5	4,67
INRAκ	2	4	3	4	3,00
MM12	6	9	6	11	7,00
INRA63	8	4	5	8	5,67
INRA005	4	4	4	6	4,00
HEL1	5	4	6	6	5,00
INRA32	7	6	5	7	6,00
INRA72	3	3	4	4	3,33
HEL5	7	5	5	8	5,67
MM8	5	5	7	7	5,67
INRA16	5	6	8	10	6,33
Moyenne	5,1	4,8	5,1	6,8	
Écart type	1,6	1,5	1,70	2,1	
Total	71	67	72	95	

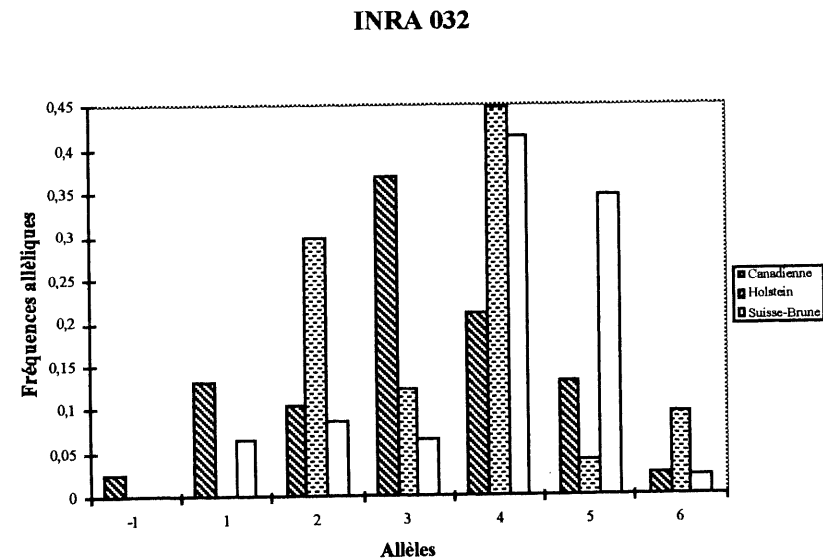
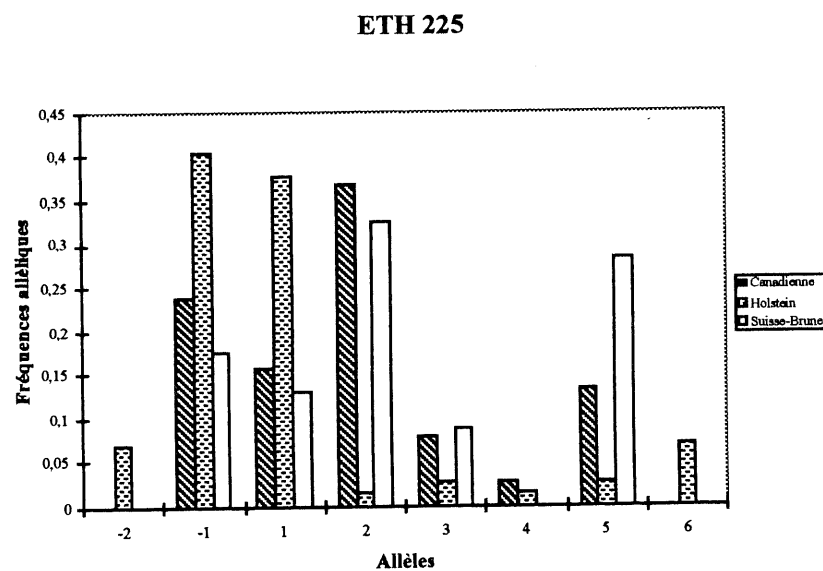
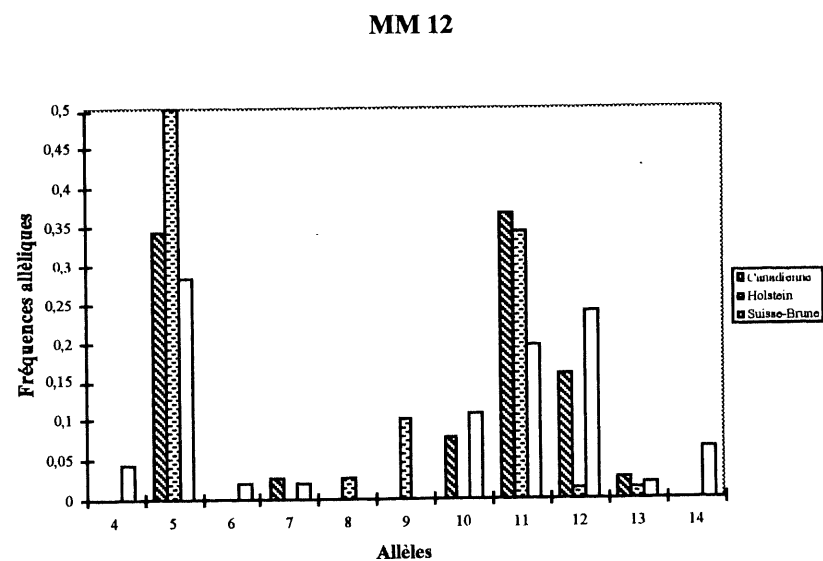
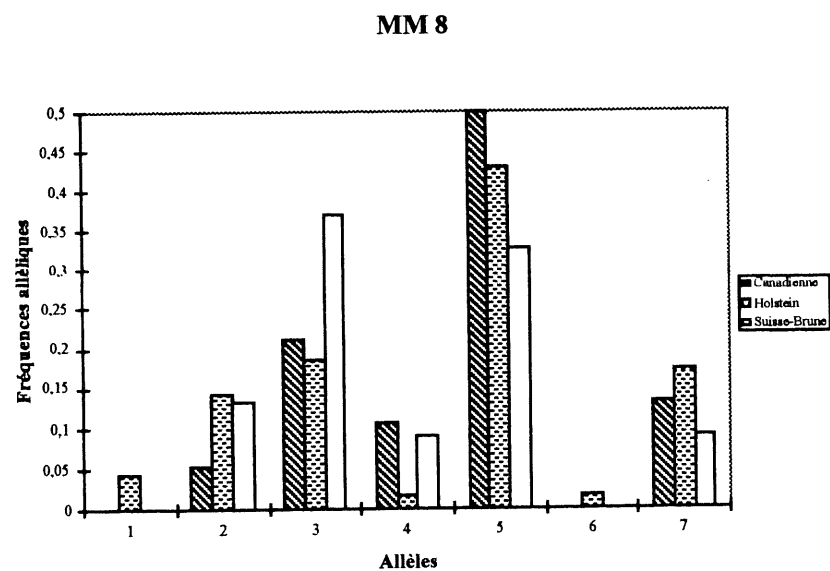


Figure 4. Exemple de distribution de fréquences alléliques à différents loci microsatellites

Tableau 8. Taux d'hétérozygotie observés et théorique non biaisé des loci microsatellites

Locus	Races							
	Canadienne		Suisse-Brune		Holstein		Moyenne par locus	
	Observé	Théorique	Observé	Théorique	Observé	Théorique	Observé	Théorique
ILSTS005	0,53	0,45	0,26	0,24	0,56	0,46	0,45	0,38
ETH225	0,63	0,78	0,78	0,78	0,61	0,70	0,68	0,75
INRA35	0,32	0,74	0,26	0,65	0,35	0,48	0,31	0,62
ETH152	0,58	0,69	0,52	0,72	0,73	0,74	0,61	0,72
INRAK	0,72	0,51	0,39	0,40	0,40	0,45	0,50	0,45
MM12	0,84	0,73	0,78	0,82	0,60	0,63	0,74	0,73
INRA63	0,84	0,76	0,52	0,48	0,53	0,56	0,63	0,60
INRA5	0,53	0,69	0,48	0,57	0,68	0,63	0,56	0,63
HEL1	0,78	0,62	0,65	0,72	0,52	0,65	0,65	0,66
INRA32	0,84	0,79	0,61	0,71	0,65	0,70	0,70	0,73
INRA72	0,89	0,58	0,43	0,50	0,68	0,62	0,67	0,57
HEL5	0,89	0,79	0,61	0,64	0,62	0,56	0,71	0,66
MM8	0,84	0,69	0,57	0,74	0,63	0,74	0,68	0,72
INRA16	0,47	0,40	0,55	0,63	0,71	0,76	0,58	0,60
Moyenne par race	0,69	0,66	0,53	0,61	0,59	0,62		

biaisé par microsatellite varie de 0,38 pour le locus ILSTS005 à 0,75 pour le locus ETH225. Parallèlement, l'hétérozygotie moyenne observée varie de 0,31 pour le locus INRA035 à 0,74 pour le locus MM12. De plus, l'hétérozygotie non biaisée moyenne est de 0,66, 0,61 et 0,62 pour les races Canadienne, Suisse-Brune et Holstein respectivement. Ces valeurs ne sont pas significativement différentes entre elles ($P > 0,1$). Notons que l'hétérozygotie observée de la race Canadienne (0,69) est significativement différente de celles des deux autres races ($P < 0,005$), soit 0,53 et 0,59 pour la race Suisse-Brune et Holstein, respectivement.

4.2.3 Équilibre de Hardy-Weinberg

L'équilibre de Hardy-Weinberg a été vérifié à différents niveaux: par locus dans chaque race, par locus pour les trois races et par race au niveau de tous les locus pris globalement. Le tableau 9 présente les valeurs de probabilité associées aux valeurs du χ^2 obtenues. L'hypothèse nulle (H_0) est que la population est en équilibre de Hardy-Weinberg tandis que l'hypothèse alternative (H_1) est que la population est en déséquilibre. Le risque d'erreur qui a été retenu est de 1 %. Donc, pour tout $P > 0,01$, on accepte H_0 et on considère que les différences entre les proportions observées et attendues sont dues au hasard. Ainsi 13 systèmes microsatellites sont en équilibre. Seul le microsatellite INRA035 présente globalement un déséquilibre par rapport à la loi de Hardy-Weinberg. Ce déséquilibre se situe principalement chez la race Canadienne et Suisse-Brune. Enfin, en considérant l'ensemble des locus microsatellites à l'intérieure de chaque race, on observe que la race Canadienne et Suisse-Brune sont globalement en déséquilibre, tandis que la race Holstein est en équilibre de Hardy-Weinberg. La race Canadienne est le plus souvent en excès d'hétérozygotes alors que la race Suisse-Brune est le plus souvent en déficit. De plus, notons que le test d'indépendance entre populations réalisé avec le programme Genepop révèle une différence significative ($P < 0,01$) entre les races au niveau de tous les loci microsatellites, à l'exception de MM8 ($P > 0,05$) (résultats non présentés). L'indépendance est également significative ($P < 0,001$) lorsque tous les loci sont pris globalement. Ainsi, 13 loci microsatellites sur 14 nous permettent de

Tableau 9. Probabilités associées au test d'équilibre de Hardy-Weinberg

Locus	Races						Les 3 races réunies P
	Canadienne		Suisse-Brune		Holstein		
	P	Écart*	P	Écart*	P	Écart*	
ILSTS005	0,025	+	1,000	+	0,405	+	0,165
ETH225	0,030	-	0,431	+	0,565	-	0,134
INRA35	0,000	-	0,000	-	0,143	-	0,000
ETH152	0,785	-	0,005	-	0,167	+	0,023
INRAK	0,151	+	1,000	+	0,466	-	0,505
MM12	0,739	+	0,528	-	0,019	-	0,133
INRA63	0,200	+	1,000	+	0,402	-	0,539
INRA5	0,143	-	0,737	-	0,383	+	0,379
HEL1	0,401	+	0,490	-	0,198	-	0,370
INRA32	0,278	+	0,013	-	0,943	-	0,076
INRA72	0,012	+	0,527	-	0,315	+	0,053
HEL5	0,124	+	0,403	-	0,492	+	0,284
MM8	0,269	+	0,009	-	0,139	-	0,013
INRA16	1,000	+	0,049	-	0,339	-	0,223
Tous les locus	0,0001		0,0002		0,1232		

*Écart de l'équilibre: (+)= excès d'hétérozygotes; (-)= déficit en hétérozygotes

différencier nos trois races entre elles. Cet aspect qui concerne la structuration des races ainsi que l'excès ou le déficit en hétérozygote, sera précisé grâce à l'analyse des F statistiques.

4.2.4 Analyse des F statistiques

La structuration de la population dans son ensemble en sous populations nous est révélée par une valeur de $F_{ST} \neq 0$. Généralement, on considère une forte structuration si $F_{ST} > 0,1$ (Wright, 1978). Ainsi, les microsatellites ILSTS005, ETH225, INRAκ, INRA072 et HEL5 ont une valeur de $F_{ST} > 0,1$ (voir Tableau 10). C'est le microsatellite ILSTS005 qui révèle la plus forte structuration avec un F_{ST} de 0,326. Ces loci nous permettent donc de bien différencier les trois races entre elles. Les autres microsatellites ont, par contre, des valeurs de $F_{ST} < 0,1$, dont MM8 qui présente la plus petite valeur (0,011). Le programme Fstat calcule également une valeur globale de F_{ST} en tenant compte de tous les loci. Cette valeur, avec un intervalle de confiance de 95 % déterminé grâce à un Bootstrap, est la suivante: $0,06 < 0,095 < 0,139$. Enfin, les tests de permutation d'allèles du programme Fstat évaluent les chances que $F_{ST} = 0$ sont inférieures à 0,4 % de façon globale. Le microsatellite MM8 présente cependant une exception à la tendance globale puisque les chances que $F_{ST} = 0$ sont de 15 %, bien que cette valeur ne soit pas jugée suffisamment élevée pour être significative.

Les valeurs de F_{IS} pour chaque race sont présentées dans le tableau 11. La valeur négative du F_{IS} de la race Canadienne suggère un excès d'hétérozygotes alors que les valeurs positives des F_{IS} des deux autres races suggèrent un déficit en hétérozygotes. Cependant, les tests de permutations du programme Fstat semblent indiquer que les chances que le F_{IS} de la race Canadienne soit égale à 0 est de 90 %, alors que cette même probabilité est de 0,4 % pour la race Suisse-Brune et de 4 % pour la race Holstein.

Tableau 10. Structuration des races

Locus	F_{st}
ILSTS005	0,326
ETH225	0,106
INRA35	0,040
ETH152	0,069
INRA κ	0,227
MM12	0,045
INRA63	0,051
INRA5	0,027
HEL1	0,045
INRA32	0,088
INRA72	0,133
HEL5	0,148
MM8	0,011
INRA16	0,071
Tous les loci considérés, avec un intervalle de confiance de 95%	$0,060 < 0,095 < 0,139$

Tableau 11. F_{is} obtenu pour chaque race

	Races		
	Canadienne	Suisse-Brune	Holstein
F_{is}	-0,050	0,139	0,047

Finalement, la valeur de F_{IT} qui a été calculée est de 0,14, ce qui indique un déficit global d'hétérozygotes de 14 % en tenant compte des trois races réunies en une même population.

4.2.5 Distances génétiques

Les distances génétiques entre les trois races sont, de façon générale, relativement petites. Dans le but de vérifier la fiabilité des noeuds composant les phénogrammes, nous avons utilisé la méthode de ré-échantillonnage du bootstrap. Cependant, cette méthode n'est vraiment valable qu'en analysant plusieurs races. C'est pourquoi, nous avons ajouté des races françaises (Moazami Goudarzi et al, 1997) et des races africaines pour ainsi parfaire l'analyse de nos trois races. Nous ne nous attarderons donc pas du tout au contexte précis de ces races supplémentaires. Notons simplement que ces races ont été analysées selon les mêmes méthodes que nos trois races avec les mêmes marqueurs microsatellites.

Deux méthodes de calcul de distance génétique (Cavalli Sforza et d standard de Nei) ont été utilisées conjointement avec deux méthodes de construction de phénogrammes (UPGMA et Neighbor Joining). Nous obtenons donc en tout quatre phénogrammes. Cependant, étant donné la grande similarité des résultats, seuls deux phénogrammes sont présentés: celui calculé à partir des distances de Cavalli Sforza et construit suivant la méthode du Neighbor Joining (voir figure 5) et celui calculé à partir des distances de Nei et construit suivant la méthode du Neighbor Joining également (voir Figure 6). Ainsi, les distances génétiques ont été calculées à partir des fréquences alléliques de 9 microsatellites (ETH225, INRA35, INRAK, INRA63, INRA005, INRA32, INRA72, INRA16 et ETH152) qui avaient également été étudiés chez les races françaises et africaines. Les arbres consensus ont été réalisés à partir de 500 bootstraps.

Les résultats nous révèlent que quelque soit la distance ou la méthode de construction de phénogramme utilisée, la race Canadienne se retrouve avec la race Suisse-Brune. Les valeurs

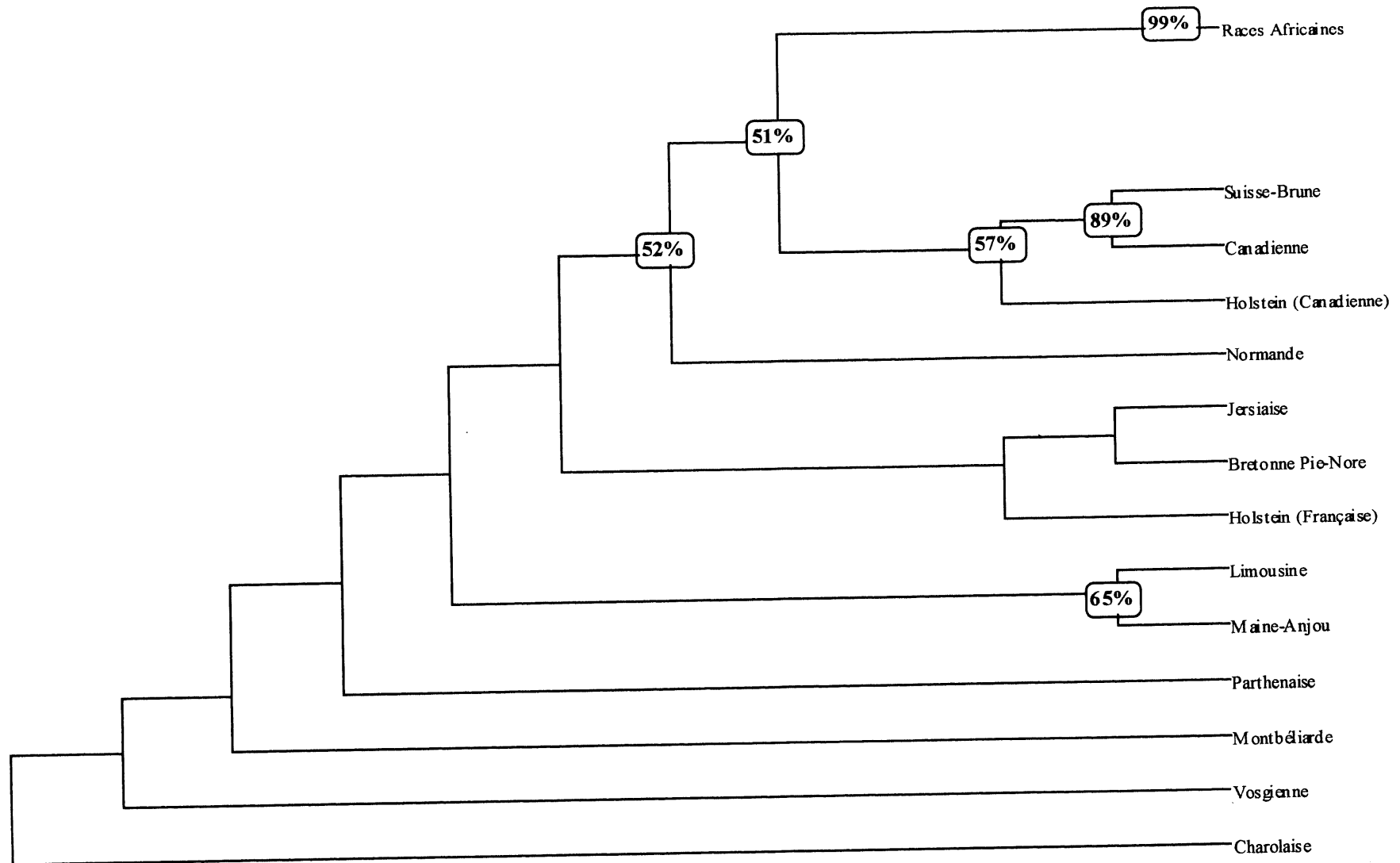


Figure 5. Phénogramme consensus des distances de Cavalli Sforza entre les races canadiennes, françaises et africaines selon la méthode du Neighbor-Joining

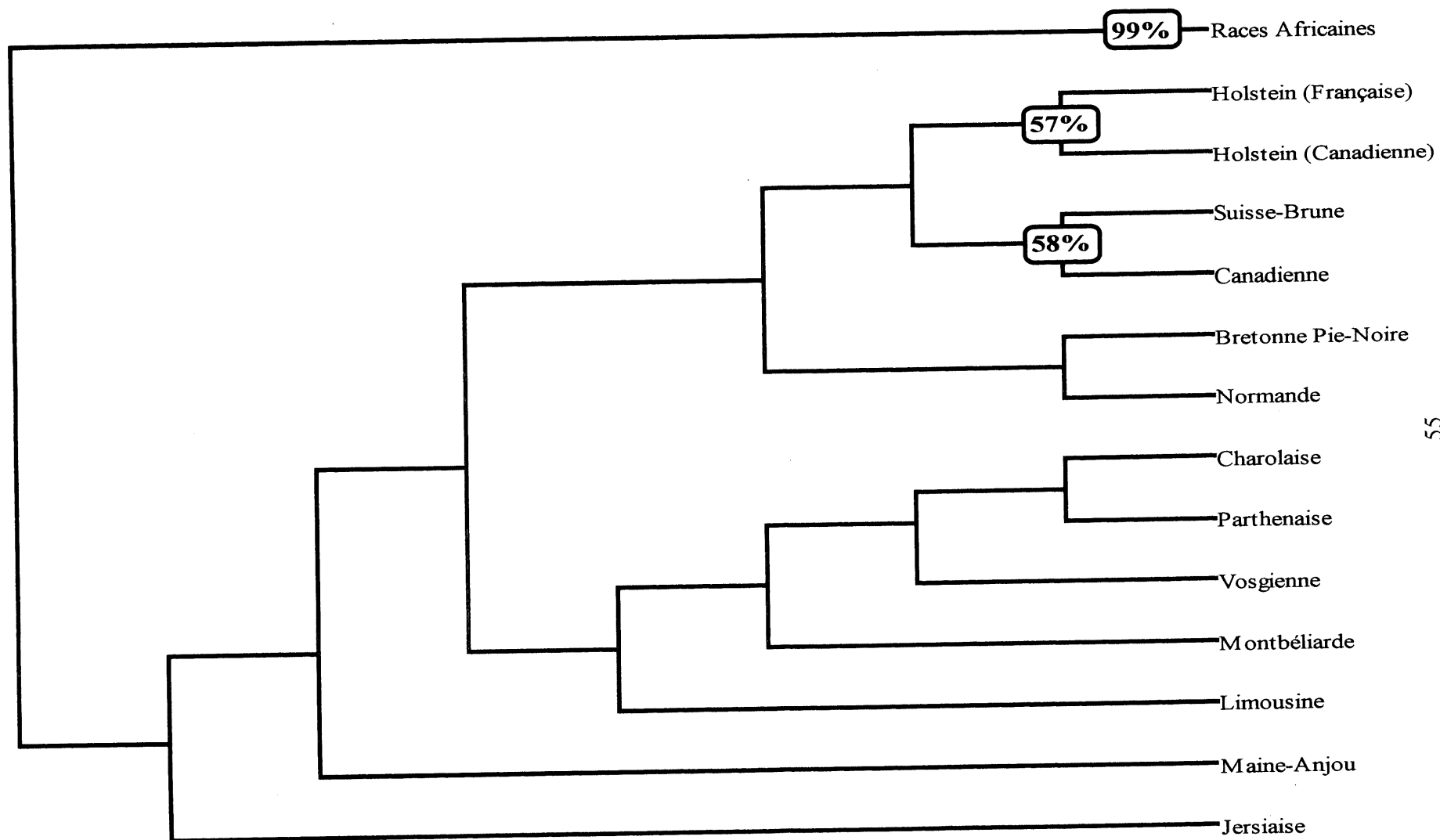


Figure 6. Phénogramme consensus des distances de Nei entre les races canadiennes, françaises et africaines selon la méthode du Neighbor-Joining

des noeuds obtenus varient de 57 à 89 %, ce qui est même supérieur aux noeuds séparant la Holstein canadienne et la Holstein française. Cependant, les races Holstein canadienne et française sont, comme on pouvait s'y attendre, généralement retrouvées ensemble (voir Figure 6). On peut également remarquer la présence de la race Normande à proximité de nos trois races (voir Figure 5) bien que les valeurs de bootstrap soient relativement faibles. D'ailleurs, la majorité des valeurs de bootstrap sont inférieures à 75 %. En fait, seul le noeud séparant les races africaines des races de bovins canadiens et français est supérieur à 95 %.

CHAPITRE 5 - DISCUSSION

Un grand nombre de marqueurs génétiques polymorphes peuvent être utilisés pour étudier la diversité génétique des populations. Le polymorphisme des protéines sériques, des protéines érythrocytaires et des lactoprotéines a souvent été utilisé pour caractériser les relations évolutives et les distances génétiques entre les races bovines (Baker et Manwell, 1980). Cependant, le polymorphisme protéique ne décrit que de façon limitée le polymorphisme réel de l'ADN. C'est pourquoi, plusieurs types de marqueurs génomiques ont été développés. Parmi les plus courants, on peut citer les marqueurs de type RFLP (restriction fragment length polymorphism), l'ADN mitochondrial, les minisatellites et les microsatellites. Lorsqu'il s'agit d'étudier les distances génétiques qui séparent des populations proches l'une de l'autre, comme des races par exemple, les marqueurs microsatellites s'avèrent être plus prometteurs parce qu'ils sont plus polymorphes. En effet, le polymorphisme mis en évidence par RFLP est souvent diallélique (Moasami-Goudarzi, 1994). De plus, la différenciation des races bovines à partir de l'ADN mitochondrial est apparemment difficile. Loftus et al. (1994), par exemple, arrivent à différencier des races bovines européennes et africaines (bovins et zébus) de races de zébus indiens, mais n'arrivent pas à différencier de façon significative les races bovines entre elles. Outre leur polymorphisme et leur nombre considérable, les microsatellites présentent des caractéristiques techniques intéressantes. Ils sont, en effet, caractérisés à l'aide de la technique de PCR ce qui en fait une procédure relativement simple et rapide. Bien que l'analyse statistique des marqueurs de type microsatellites reste complexe, de plus en plus de chercheurs s'y intéressent (Buchanan et al., 1994; MacHugh et al., 1994, 1997; Moazamie Goudarzy et al., 1997).

5.1 Polymorphisme des caséines

Les caséines sont des protéines d'une importance économique certaine pour les bovins laitiers, notamment au niveau d'éventuelles relations entre leur polymorphisme et la production de lait, sa composition et ses aptitudes fromagères (Grosclaude, 1988). Alors que la plupart des races bovines sont bien caractérisées pour ces loci, très peu d'études ont été réalisées chez la race Canadienne. Notre étude nous permet donc de dresser un portrait du polymorphisme de ces loci chez la race Canadienne. Ainsi, nous observons que les fréquences alléliques obtenues au sein de l'échantillon de la race Canadienne ressemblent

d'avantage aux fréquences alléliques obtenues pour la race Holstein. Cependant, le polymorphisme caséique que nous détectons étant relativement limité, le polymorphisme des microsatellites sera privilégié pour ce genre d'analyse. La même remarque s'applique aux indices de variabilité tel que le nombre moyen d'allèles par locus et le taux d'hétérozygotes, qui sont tout deux moins élevés qu'au niveau des marqueurs microsatellites.

Notons que de façon générale, les résultats obtenus chez nos trois races concordent avec les résultats disponibles dans la littérature (Ng-Kwai-Gang, 1992; Swaisgood, 1992). Seule la Suisse-Brune présente un certain écart. Les différences entre les fréquences alléliques que nous avons obtenues et celles obtenues par les auteurs consultés peuvent en partie s'expliquer par le nombre relativement restreint d'animaux analysés dans notre étude et par des critères d'échantillonnage différents. En effet, les auteurs consultés ont étudié un grand nombre d'animaux apparentés entre eux. Sans doute aurait-il été intéressant pour notre étude d'analyser un plus grand nombre d'animaux pour en arriver à des résultats plus proches des valeurs réelles de la population. Cependant, le fait que nous ayons choisi des individus non apparentés, du moins jusqu'à la deuxième génération, nous permet d'obtenir une image plus globale du polymorphisme des caséines au sein de nos trois populations. De plus, en ce qui concerne la race Canadienne, les individus étudiés représentent probablement une grande partie des familles existant dans cette race, étant donné son faible effectif.

5.2 Polymorphisme des microsatellites

Les microsatellites se sont vite révélés comme étant des outils intéressants pour l'étude de la génétique des populations. L'utilisation croissante de ces marqueurs tend à le démontrer.

5.2.1 Variabilité et structure des populations

Le premier objectif de mon travail consistait à évaluer la variabilité génétique de trois races bovines, dont en particulier la race Canadienne. Ainsi, les microsatellites analysés se révèlent être passablement polymorphes étant donné une moyenne de 6,8 allèles détectés par loci. Cette moyenne est satisfaisante puisqu'elle correspond aux données de la littérature (Vaiman et al., 1994). La race Canadienne présente un nombre moyen d'allèles par locus comparable à ceux des races Holstein et Suisse-Brune. Les taux d'hétérozygoties sont généralement élevés pour l'ensemble des trois races. L'hétérozygotie observée est significativement plus élevée

chez la race Canadienne que chez la Suisse-Brune et la Holstein. Cependant, l'écart entre les taux d'hétérozygotie non biaisés des trois races n'est pas très important, bien que ce taux soit légèrement plus élevé pour la race Canadienne. L'analyse de l'équilibre de Hardy-Weinberg et des F statistiques nous apporte des précisions intéressantes à ce sujet. Ainsi, seul le microsatellite INRA035 est largement en déséquilibre par rapport à la loi de Hardy-Weinberg, principalement chez la race Canadienne et Suisse-Brune. Ces deux races présentent d'ailleurs un écart significatif à l'équilibre de Hardy-Weinberg, contrairement à la race Holstein qui est globalement à l'équilibre. Elles diffèrent cependant quand à l'origine de cet écart. En effet, la race Canadienne a le plus souvent un excès (significatif ou non) d'hétérozygotes tandis que la race Suisse-Brune présente une majorité de locus en déficit d'hétérozygotes (significatif ou non). De plus, la valeur négative du F_{IS} de la race Canadienne (-0,05), qui est l'équivalent au taux de consanguinité, semble confirmer son excès d'hétérozygotes. De la même façon, le F_{IS} de la race Suisse-Brune confirme son déficit en hétérozygotes étant donné sa valeur positive (0,139). Cependant, les tests de permutations effectués avec le programme Fstat estime la probabilité, chez la race Canadienne, que $F_{IS} = 0$ légèrement supérieure à 90%. Précisons que ce test de permutation n'est utile que si on l'effectue sur un fichier ne contenant qu'une seule population à la fois. Fstat crée alors une population "imaginaire" pour réaliser les calculs. Il est fort probable que la valeur du F_{IS} de la race Canadienne ne soit pas significativement différente de zéro. Le même test effectué séparément sur les deux autres races révèle une valeur significativement supérieure à zéro, bien que de façon moins évidente pour la Holstein puisque cette race est plutôt à l'équilibre. D'ailleurs, il est intéressant de remarquer que, d'après une étude de généalogie, Miglior et Burnside (1995) n'observent pas une consanguinité très élevée chez la race Holstein.

Le déficit en hétérozygotes observé chez la Suisse-Brune s'explique plus facilement que l'excès de la race Canadienne. Plusieurs facteurs pourraient expliquer le déséquilibre observé: mis à part le hasard, en ce qui concerne les causes d'origine génétique, il est bien connu que la consanguinité modifie les fréquences génotypiques. La conséquence en est une perte de variabilité génétique de la population au cours des générations. Ainsi, une partie des déséquilibres observés s'expliqueraient par la consanguinité des animaux choisis au sein de la race Suisse-Brune. Un des autres facteurs pourrait être l'existence d'allèles nuls, allèles ne donnant lieu par PCR à aucune amplification. Une délétion aux niveau des amorces ou une mutation dans les séquences flanquantes du microsatellite peuvent entraîner la présence

d'allèles nuls. Parmi les 14 microsatellites utilisés dans notre étude, 9 sont communs à l'étude de Moazami-Goudarzi (1997). Sur les 9 loci, après analyse de ségrégation sur plusieurs familles, aucun allèle nul n'a pu être détecté sauf pour INRA035 dans quelques cas. Par ailleurs, les races étudiées étant des races sélectionnées, il n'est pas à exclure que certains des microsatellites choisis soient proches de gènes sélectionnés, situation pouvant conduire à un écart à la loi de Hardy-Weinberg.

Pour ce qui est de l'excès en hétérozygote de la Canadienne, on pourrait imaginer que l'apport génétique de la Suisse-Brune suite aux croisements effectués dans les années 70, ait eu un effet sur le taux d'hétérozygotes ou bien que les calculs ne soient pas assez exacts du fait du faible nombre d'animaux testés. En fait, le nombre d'individus constituant l'échantillon est malheureusement faible ($n = 19$). Il n'a pas été possible de trouver plus de 20 individus qui respectaient la règle que nous nous étions fixée (non apparenté jusqu'à la deuxième génération). Quoi qu'il en soit, le F_{IS} ne révèle pas une consanguinité plus élevée pour la race Canadienne que pour la race Holstein. Evidemment, nous devons également nous soucier de la variation de cette consanguinité dans le temps avant d'émettre des conclusions. De plus, seulement 24 taureaux sont disponibles en moyenne au sein de la race Canadienne (Hansen, 1997). Il est donc important de suivre l'évolution de la consanguinité, aussi bien chez la race Canadienne qu'au niveau des deux autres races.

5.2.2 Distances génétiques

La seconde partie de mon travail consistait à évaluer les relations génétiques entre les trois races d'intérêt. Avant de calculer les distances génétiques, certains tests statistiques peuvent déjà nous renseigner sur le potentiel de nos marqueurs microsatellites à différencier les races étudiées. Le premier test est celui du χ^2 qui compare les distributions des fréquences alléliques des différents locus d'une race à l'autre. Les résultats de ce test indiquent que les fréquences alléliques sont significativement ($P < 0,05$) différentes d'une race à l'autre. De plus, le logiciel Genepop permet également de tester l'indépendance de nos trois races au niveau de la distribution de leurs fréquences alléliques. Les valeurs de P obtenues sont toutes inférieures à 1 % ce qui confirme l'indépendance des populations, à l'exception du locus MM8 où cette valeur est de 9 %. Le microsatellite MM8 différencie donc moins efficacement les races entre elles. Enfin, le coefficient F_{ST} confirme la structuration de l'ensemble des

individus étudiés en trois races. En effet, selon le barème que Wright (1978) propose, la valeur de F_{ST} obtenue, soit 0,095, est considérée comme présentant une structuration modérée. De plus, selon les tests de permutations d'allèles réalisés à l'aide du programme Fstat, les chances pour que $F_{ST} = 0$ sont très faibles ($< 1\%$). Nos marqueurs microsatellites peuvent donc bien différencier nos trois races entre elles.

Les distances génétiques obtenues (D_s de Nei et D_c de Cavalli Sforza) entre les trois races d'intérêt sont cependant relativement petites. Dans le but de valider ces résultats, nous les avons intégrés aux données obtenues par Moazamie-Goudarzi et al. (1997). En tout, 9 microsatellites sont communs aux deux études. Les données de Moazamie-Goudarzi portent sur 10 races bovines françaises (Moazamie Goudarzi et al., 1997) ainsi que sur quatre races africaines (non publié). De ces analyses, deux principales conclusions peuvent être amenées. Premièrement, la race Canadienne se retrouve toujours avec la Suisse-Brune peu importe le type de distance et la méthode de construction d'arbre. Les regroupements (noeuds) obtenus entre ces deux races à l'intérieur de l'arbre consensus sont relativement forts (57 à 80%) si on les compare aux autres noeuds obtenus. Ils sont même quelques fois plus forts que les noeuds entre la Holstein canadienne et française. Deuxièmement, nous retrouvons globalement les races Holstein canadienne et française très près l'une de l'autre, ce qui nous rassure quant à la validité de notre approche. De plus, on peut observer que les trois races canadiennes se regroupent ensemble par rapport aux races françaises. En fait, aucune race française ne semble vraiment s'apparenter à l'une de nos trois races. Pourtant, on considère souvent que la race Canadienne et la race Jersey (qui a été étudiée par Moazamie-Goudarzi) sont apparentées. Rien ne nous l'indique dans nos résultats.

Cependant, il est important de réaliser que de façon générale, les valeurs des noeuds sont très faibles (souvent $< 75\%$) comparées au seuil de 95% que l'on se fixe habituellement. En fait, seul le branchement entre les races africaines et l'ensemble des autres races présente une valeur de bootstrap $> 95\%$. Ces faibles valeurs de bootstrap peuvent s'expliquer soit par une sous-estimation de la mesure de la distance génétique soit par un nombre insuffisant de microsatellites (Moazamie-Goudarzi et al., 1997). En effet, il est possible que les méthodes de calcul de distances génétiques que nous avons utilisées entraînent une sous-estimation de ces distances étant donné leurs propriétés mathématiques. En fait, certains auteurs considèrent que la distance de Nei, par exemple, s'applique moins bien aux microsatellites

parce qu'elle est basé sur le modèle IAM (infinite allele model), alors que les microsatellites semblent plutôt suivre le modèle SMM (stepwise mutation model) (Shriver et al., 1993; Valdes et al., 1993). Cependant, Takezaki et Nei (1996) ont démontré que les changements évolutifs des loci microsatellites ne sont pas importants lors de la reconstruction phylogénétique lorsque les populations étudiées sont intimement liées (des races, par exemple). Enfin, l'hypothèse la plus probable reste que le nombre de loci microsatellites soit insuffisant. En effet, plus les races sont proches plus le nombre de loci nécessaires pour les séparer est grand. Bien que, en général, les valeurs des bootstraps augmentent avec le nombre de loci considérés, Moazamie-Goudarzi et al. (1997) n'observent pas d'évolutions significatives des valeurs de bootstrap calculées avec 5 ou 17 microsatellites. À la lumière de ces résultats, ces auteurs estiment que le nombre de microsatellites nécessaires pour discriminer des races proches soit supérieur à 30, un nombre pourtant considéré généralement comme suffisant (Moazamie-Goudarzi et al., 1997).

CHAPITRE 6 - CONCLUSION

En analysant le polymorphisme des caséines et des marqueurs microsatellites, nous avons pu évaluer la variabilité génétique des races Canadienne, Suisse-Brune et Holstein. Initialement, nous avons émis l'hypothèse que la race Canadienne était moins variable et présentait une consanguinité plus importante que la race Holstein. Contrairement à ce que nous supposions au départ, la race Canadienne ne semble pas être moins variable que les deux autres puisqu'elle présente un nombre d'allèles par locus et un taux d'hétérozygotie comparables sinon supérieurs à ceux des deux autres races. La race Canadienne et Suisse-Brune ne vérifient pas les proportions de Hardy-Weinberg. Pour la Canadienne, cet écart s'observe en raison d'un excès en hétérozygotes, tandis que la Suisse-Brune présente un déficit en hétérozygotes. Ainsi, la race Suisse-Brune semble présenter une consanguinité plus importante que les deux autres. De plus, l'apport génétique des croisements de la race Canadienne avec la Suisse-Brune s'observe au niveau des phénogrammes illustrant les distances génétiques. En effet, la Canadienne semble être plus apparentée à la Suisse-Brune, ce qui confirme notre seconde hypothèse. Ces deux races semblent cependant toutes deux apparentées à la Holstein. Bien que cette étude présente des résultats intéressants, un nombre plus grand de microsatellites (ainsi qu'un échantillonnage plus important) serait nécessaire pour confirmer nos résultats. Cependant, de telles études demandent une quantité considérable de ressources. C'est pourquoi il devient important pour ce genre d'étude d'établir un consensus entre chercheurs sur les microsatellites à utiliser. Le type d'analyse statistique demande également qu'on s'y attarde davantage puisqu'il demeure difficile de porter un jugement éclairé concernant la conservation de telle ou telle race.

BIBLIOGRAPHIE

- ASCHAFFENBURG, R. 1961. Inherited casein variants in cow's milk. *Nature*. 192: 431-432.
- ASCHAFFENBURG, R. et J. DREWRY. 1955. Occurrence of different beta-lactoglobulins in cows milk. *Nature*. 176: 218-219.
- ASCHAFFENBURG, R., A. SEN et M.P. THOMPSON. 1968. Genetic variants of casein in Indian and African Zebu cattle. *Comp. Biochem. Physiol.* 25: 177-184.
- BAKER, C.M.A. et C. MANWELL 1980. Chemical classification of cattle. 1. Breed groups. *Anim. Bolld Groups Biochem. Genet.* 11: 127-150.
- BELL, K. 1962. One-dimensional starch gel electrophoresis of bovine skim milk. *Nature*. 195: 705-706.
- BELL, K., H. A. MCKENZIE et D.C. SHAW. 1981b. Bovine b-lactoglobulin E, F, and G of Bali (Banteng) cattle, *Bos (Bibos) Javanicus*. *Aust. J. Biol. Sci.* 34: 133-147.
- BELL, K., H. A. MCKENZIE, U. MULLER, C. ROGERS et D.C. SHAW. 1981a. Equine whey protein. *Comp. Biochem. Physiol.* 68B: 225-236.
- BELL, K., K. E. HOPPER et H.A. MCKENZIE. 1981c. Bovine a-lactalbumin C and as1-, b- and k-casein of Bali (Banteng) cattle, *Bos (Bibos) javanicus*. *Aust. J. Biol. Sci.* 34: 149-159.
- BERNIER, J. G. 1995. Canadienne bovine breed report. *Proc. Third Global Conf. on Conserv. Domest. Anim. Genet. Resour., Rare Breeds International*, Kenilworth, U.K. 1488-156.
- BISHOP, M.D., S. M. KAPPES, J.W. KEELE, R.T. STONE, S.L.F. SUNDEN, G.A. HAWKINS, S.S. TOLDO, R. FRIES, M.D. GROSZ, J. YOO et C.W. BEATTIE. 1994. A genetic linkage map for cattle. *Genetics* 136: 619-639.
- BLUMBERG, B.S. et M. P. TOMBS. 1958. Possible polymorphism of bovine a-lactalbumin. *Nature*. 181: 683-684.
- BOWCOCK A.M., A.R. LINARES, J. TOMFOHRDE, E. MINCH, J.R. KIDD AND L.L. CAVALLI-SFORZA. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*. 368: 455-457.
- BREZENSKY, L., S. J. KEMP, A. J. TEALE. 1993. ILSTS 005: a polymorphic bovine microsatellite. *Anim. Genet.* 24: 73.

- BUCHANAN, F.C., LINDA J. A., ROGER P.L., MADDOX J.F., CRAWFORD A.M. 1994. Determination of Evolutionary relationships among sheep breeds using microsatellites. *Genomics*. 22: 397-403.
- CAVALLI SFORZA, L. L. et W. F. BODMER. 1971. *Genetics, Evolution and Man*. Freeman, San Francisco.
- CAVALLI SFORZA, L. et W. EDWARDS. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution*. 21: 550-570.
- CLRC. 1995. Annual Report. Canadian Livestock Records Corporation, Ottawa, Canada, 68pp.
- COUTURE, J.A. 1909. *Le Bétail Canadien. Livre de généalogie du bétail canadien*. 1: 5-12.
- FALCONER, D.S. 1989 *Introduction to quantitative genetics* 3rd. edition. Longmann Scientific and Technical, U.K.: 438pp.
- F.A.O., 1993. An integrated global programme to establish the genetic relationships among the breeds of each domestic animal species. Report of Animal Production and Health Division.
- FELSENSTEIN, J. 1993. PHYLIP Phylogeny inference Package, Version 3.5 Edition. Department of Genetics, Washington University, Seattle..
- FELSENSTEIN, J. 1995. Confidence limits on phylogenies : an approach using the bootstrap. *Evolution* 39: 783-791.
- FITCH, W. M. et E. MARGOLIACH. 1967. Construction of phylogenetic tress. *Science*. 155: 279-284.
- FORTIN, L. DE G. 1940 *Histoire de la race bovine Canadienne*. La Bonne Terre, École supérieure d'Agriculture, St-Anne de la Pocatière, Québec, Canada.: 286pp.
- FSTAT, 1984. A program to calculate Weir and Cockerham's estimation of F-statistic (version 1.0).
- GREGORIUS, H.-R. 1984. A unique genetic distance . *Biometrical Journal*. 26: 13-18.
- GROSCLAUDE, F. 1988. Le polymorphisme génétique des principales lactoprotéines bovines. *INRA Prod. Anim.*. 1: 5-17.
- GROSCLAUDE, F., J. PUJOLLE , J. GARNIER et B. RIBADEAU-DUMAS. 1965. Déterminisme génétique des caséines k du lait de vache; étroite liaison du locus k-Cn avec les loci as1-Cn et b-Cn. *C.R. Hebd. Séances Acad. Sci., Paris*. 261: 5229-5232.

- GROSCLAUDE, F., J. GARNIER, B. RIBADEAU-DUMAS et R. JEUNET. 1964. Étroite dépendance des loci contrôlant le polymorphisme des caséines as et b. C.R. Hebd. Séance Acad. Sci., Paris. 259: 1569-1571.
- GROSCLAUDE, F., J. PUJOLLE, J. GARNIER et B. RIBADEAU-DUMAS. 1966. Mise en évidence de deux variants supplémentaires des protéines du lait de vache: as1-Cn D et Lg D. Ann. Biol. Anim. Biochim. Biophys. 6: 215-222.
- GROSCLAUDE, F., M. F. MAHÉ, J.C. MERCIER, J. BONNEMAIRE et J.H. TEISSIER. 1976. Polymorphisme des lactoprotéines de bovins Néplais. II. Polymorphisme des caséines 'as-mineurs'; le locus as2-Cn est-il lié aux loci as1-Cn, b-Cn et kCn? Ann. Génét. Sél. Anim. 8: 481-491.
- GROSCLAUDE, F., P. JOUDRIER et M.F. MAHÉ. 1979. A genetic and biochemical analysis of polymorphism of bovine as2-casein. J. Dairy Res. 46: 211-213.
- GROSCLAUDE, F., P. JOUDRIER et M.F. MAHÉ. 1978. Polymorphisme de la caséine as2 bovine: étroite liaison du locus as2-Cn avec les loci a-Cn, b-Cn et k-Cn; mise en évidence d'une délétion dans le variant as2-Cn D. Ann. Génét. Sél. Anim. 10: 313-327.
- HALIASSOS, A., J. C. CHOMEL, L. TESSON, M. BAUDIS, J. KRUH, J. C. KAPLAN et A. KITZIS. 1989. Modification of enzymatically amplified DNA for the detection of point mutations. Nucl. Acids Res. 17: 3606.
- HAMADA, H., M. G. PETRINO et T. KAKUNAGA. 1982. A novel repeated element with Z-DNA forming potential is wide found in evolutionarily diverse eukaryotic genomes. Proc. Natl. Acad. Sci. 79: 6465-6469.
- HANSEN, C. 1997. The genetic distance among cattle breeds as related to animal genetic resource conservation. Thèse de Doctorat, University of Manitoba.
- HARTL, D.L. 1988 Génétique des populations. Médecine-Science Flammarion, Paris.: pp.305.
- HUDSON, G.F.S. et L.D. VAN VLECK. 1984. Effects of inbreeding on milk and fat production, stayability, and calving interval of registered Ayrshire cattle in Northeastern United States. J. Dairy. Sci. 67: 171-179.
- KAUKINEN, J. et S. L. VARVIO. 1993. Eight polymorphic bovine microsatellites. Anim. Genet.. 24: 148.
- KIDDY, C.A., R. E. MCCANN et W.W. THATCHER. 1968. Gene frequencies in milk protein polymorphisms in dairy cattle. Immunogenetics Letter. 5: 150-152.
- LEVEZIEL, H. C. RODELLAR, C. LEROUX, L. PEPIN, C. GROHS, D. VAIMAN, M.F. MAHÉ, P. MARTIN et F. GROSCLAUDE. 1994. A microsatellite within the bovine k-casein gene reveals a polymorphism correlating strongly with polymorphisms previously described at the protein as well as the DNA level. Anim. Genet. 25: 223-228.

- LI, F.H.F. et S.N. GAUNT. 1972. Study of genetic polymorphisms of milk b-lactoglobulin, α s1-casein, b-casein and k-casein in five dairy breeds. *Bioch. Genet.* 6: 9-20.
- LIEN, S., P. ALESTRÖM, H. KLUNGL et S. ROGNE. 1992. Detection of multiple b-casein (CASB) alleles by Amplification Created Restriction Sites (ACRS). *Anim. Genet.* 23: 333-338.
- LIEN, S., S. KAMINSKI, P. ALESTRÖM et A. ROGNE. 1993. A simple and powerful method for linkage analysis by amplification of DNA from sperm cells. *Genomics.* 16: 41-44.
- LOFTUS, R. T., D.E. MACHUGH, L.O. NGERE, D.S. BALAIN, A.M. BADI, D.G. BRADLEY et E. P. CUNNINGHAM. 1994. Mitochondrial genetic variation in European, African and Italian cattle populations. *Anim. Genet.* 25: 265-271.
- LOUIS, E. J. et E. R. DEMPSTER. 1987. An exact test for Hardy-Weinberg and multiple alleles. *Biometrics.* 43: 805-811.
- MACHUGH, D. E., M. D. SHRIVER, R. T. LOFTUS, P. CUNNINGHAM et D. G. BRADLEY. 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of Taurine and Zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics.* 146: 1071-1086.
- MACHUGH, D. E., R. T. LOFTUS, D. G. BRADLEY, P. M. SHARP et P. CUNNINGHAM. 1994. Microsatellite DNA variation within and among European cattle breeds. *Proc. R. Soc. Lond.* 256: 25-31.
- MARTIN, P. et F. GROSCLAUDE. 1993. Improvement of milk protein quality by gene technology. *Livestock Production Science.* 35: 95-115.
- MERCIER, J.C. GROSCLAUDE F. 1993. Génétique moléculaire des protéines du lait et de leurs gènes. *Biologie de la lactation.* Edition INSERM/INRA. 319-347.
- MIGLIOR, F. et E.B. BURNSIDE. 1995. Inbreeding of Canadian Holstein cattle. *J. Dairy Sci.* 78: 1163-1167.
- MOAZAMI-GOUDARZI, K., D. LALOË, J. P. FURET et F. GROSCLAUDE. 1997. Analysis of genetic relationships between 10 cattle breeds with 17 microsatellites. *Anim. Genet.* 28: 338-345.
- MOMMENS, G.W., A. COPPIETEN, A. VAN DE WEGHE, A. VAN ZEVEEREN et Y. BOUQUET. 1994. Dinucleotide repeat polymorphism at the bovin MM12EG and MM8D3 loci. *Anim. Genet.* 25: 368.
- NEI, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics.* 89: 583-590.

- NEI, M. 1972. Genetic distance between populations. *The American Naturalist*. 106: 283-292.
- NEI, M. et A.K. ROYCHOUDHURY. 1974. Sampling variances of heterozygosity and genetic distance. *Genetics*. 76: 379-390.
- NG-KWAI-HANG, K.F. et F. GROSCLAUDE. 1992, Genetic Polymorphism of milk proteins in *Advanced Dairy Chemistry Volume 1*. Elsevier Science Publishers LTD.: 405-456.
- NG-KWAI-HANG, K.F., J. F. HAYES, J.E. MOXLEY et H.G. MONARDES. 1986. Relationships Between Milk Protein polymorphisms and major milk constituents in Holstein-Friesian Cows. *J. Dairy Sci.* 69: 22-26.
- NG-KWAI-HANG, K.F., MONARDES H. G. et J.F. HAYES. 1990, Association between genetic polymorphism of milk proteins and production traits during three lactations. *J. of dairy Sci.* 73: 3414-3420.
- PETERSON, R.F. et F.C. KOPFLER. 1966. Detection of new types of b-casein by polyacrylamide gel electrophoresis at acid pH: a proposed nomenclature. *Biochem. Biophys. Res. Commun.* 22: 388-392.
- QUELLAR, D. C., J. E. STRASSMANN et C. R. HUGHES. 1993. Microsatellites and kinship. *Trends Ecol. Evol.* 8: 285-288.
- RAYMOND, M. et F. ROUSSET. 1994. An exact test for population differentiation. .
- RAYMOND, M et ROUSSET F. 1995. GENEPOP (version 1,2): Population genetic software for exact test and ecumenicism. *J. Heredity*. 86: 248-249.
- RIENZO, A. DI, A. C. PETERSON, J.C. GARZA, A. M. VALDES, M. SLATKIN et N. A. FREIMER. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Genetics*. 91: 3166-3170.
- SAITOU, N. et NEI M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.
- SAS INSTITUTE. 1989 SAS User's Guide: Statistics, (Version 6 Edition by SAS Institute), . Cary, NC.: 846pp.
- SCHLÖTTERER, C. et D. TAUTZ. 1992. Slippage synthesis of simple sequence DNA. *Nucl. Acids Res.* 20: 211-215.
- SEIBERT, B., G. ERHARDT et B. SENFT. 1985. Procedure for simultaneous phenotyping of genetic variants in cow's milk by isoelectric focusing. *Animal Blood Group and Biochemical Genetics*. 16: 183-191.

- SHRIVER, MARK D., LI JIN , RANAJIT CHAKRABORTY et ERIC BOERWINKLE. 1993. VNTR allele frequency distributions under the Stepwise Mutation Model: A computer Simulation Approach. *Genetics*. 134: 983-993.
- SIMON, D.L. 1984. Conservation of animal genetic resources - A review. *Livest. Prod. Sci.* 11: 23-36.
- SINGER, M.F. 1982. SINEs and LINEs: Highly repeated short and long interspersed sequences in mammalian genome. *Cell*. 28:433-434.
- SNEATH, P.H.A. et R.R. SOKAL. 1973. Numerical taxonomy. W.H. Freeman and Company, San Francisco. 573 pp.
- SOKAL, R.R. et C.D. MICHENER. 1958. A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull.* 38: 1409-1438.
- SOLIGNAC, M., G. PERIQUET , D. ANXOLABÉHÈRE et C. PETIT. 1995. Génétique et évolution Tome 1. La variation, les gènes dans les populations. Collection Méthodes; Hermann, Éditeurs des sciences et des arts. 295 pp.
- STALLINGS, R.L., A. F. FORD, D. NELSON, D.C. TORNEY, C.E. HILDEBR et R.K. MOYZIS. 1991. Evolution and distribution of (GT)_n repetitive sequences in mammalian genomes. *Genomics*. 10: 807-815.
- STEFFEN, P., A. EGGEN , A. B. DIETZ, J. E. WOMACK, G. STRANZINGER et R. FRIES. 1993. Isolation and mapping of polymorphic microsatellites in cattle. *Anim. Genet.* 24: 121-124.
- SWAISGOOD, H.E. 1992. Chemistry of the caseins: in *Advanced Dairy Chemistry Vol. 1 Protein*. Elsevier Science Publishers LTD, London. 63-110.
- TAKEZAKI N. et M. NEL. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*. 144: 389-399.
- TAUTZ, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucl. Acids Res.* 17: 6463-6471.
- THOMPSON, M.P., C. A. KIDDY, L. PEPPER et C.A. ZITTLE. 1962. Variation of the as-casein fraction of individual cow's milk. *Nature*. 195: 1001-1002.
- TREADGILL, W.D. et J.E. WOMACK. 1990. Genomic analysis of the major bovine casein genes. *Nucl. Acid Res.* 18: 6935-6942.
- VAIMAN, D., D. MERCIER , K. MOAZAMI-GOUDARZI, A. EGGEN, R. CIAMPOLINI, A. LÉPINGLE, R. VELMALA, J. KAUKINEN, S. L. VARVIO, P. MARTIN et H. LEVESIEL. 1994. A set of 99 cattle microsatellites: characterization, syntenic mapping, and polymorphism. *Mamm Genome*. 5 : 288-297.

- VALDES, A.M., M. SLATKIN et N.B. FREIMER. 1993. Allele frequencies at microsatellite loci - the stepwise mutation model revisited. *Genetics*. 133: 737-749.
- VASSART, M. 1995. Nouveaux outils en génétique de la conservation. *Revue Méd.* 146: 171-180.
- VOGLINO, G.F. 1972. A new b-casein variant in Piedmont cattle. *Anim. Bolld. Grps Biochem. Genet.* 3: 61-62.
- WAHLUND, . 1928. *Hereditas* II. 65.
- WEBER, J. L. et MAY, P. E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* 44: 388-396.
- WRIGHT, S. 1978. *Evolution and the genetic of population*. University of Chicago Press, Chicago. 295p.
- WU, C.F.J. 1986. Jackknife, bootstrap and other resampling plan in regression analysis. *Annals Statistic.* 14: 1261-1295.

Annexe 1. Fréquences alléliques des loci microsatellites

ILSTS 005

	Allèles					
	1	3	5	6	8	9
Canadienne	0,0000	0,0000	0,7105	0,2368	0,0263	0,0263
Holstein	0,0000	0,0139	0,3194	0,6667	0,0000	0,0000
Suisse-Brune	0,0217	0,0000	0,8696	0,1087	0,0000	0,0000

ETH 225

	Allèles							
	-2	-1	1	2	3	4	5	6
Canadienne	0,0000	0,2368	0,1579	0,3684	0,0789	0,0263	0,1316	0,0000
Holstein	0,0694	0,4028	0,3750	0,0139	0,0278	0,0139	0,0278	0,0694
Suisse-Brune	0,0000	0,1739	0,1304	0,3261	0,087	0,0000	0,2826	0,0000

INRA 35

	Allèles				
	3	5	7	8	9
Canadienne	0,1579	0,1579	0,0789	0,4474	0,1579
Holstein	0,0135	0,0000	0,0000	0,6351	0,3514
Suisse-Brune	0,0435	0,2174	0,0000	0,5217	0,2174

ETH 152

	Allèles				
	1	2	3	4	5
Canadienne	0,0526	0,2895	0,0000	0,2105	0,4474
Holstein	0,1487	0,1487	0,1757	0,4324	0,0946
Suisse-Brune	0,0247	0,3261	0,0435	0,3261	0,2826

INRA κ

	Allèles			
	3	4	6	7
Canadienne	0,0000	0,4722	0,5278	0,0000
Holstein	0,0000	0,6857	0,2857	0,0286
Suisse-Brune	0,0435	0,1739	0,7609	0,0217

MM12

MM112	Allèles											
	4	5	6	7	8	9	10	11	12	13	14	
Canadienne	0,0000	0,3421	0,0000	0,0263	0,0000	0,0000	0,0790	0,3684	0,1579	0,0263	0,0000	
Holstein	0,0000	0,5000	0,0000	0,0000	0,0286	0,1000	0,0000	0,3429	0,0143	0,0143	0,0000	
Suisse-Brune	0,0435	0,2826	0,0217	0,0217	0,0000	0,0000	0,1087	0,1957	0,2391	0,0217	0,0652	

INRA 63

	Allèles							
	-1	1	3	4	5	6	7	9
Canadienne	0,0526	0,0263	0,1842	0,3158	0,3421	0,0263	0,0263	0,0263
Holstein	0,0139	0,0139	0,0417	0,4028	0,5278	0,0000	0,0000	0,0000
Suisse-Brune	0,0000	0,0435	0,0435	0,2174	0,6957	0,0000	0,0000	0,0000

INRA 005

	Allèles					
	-4	-1	1	2	3	7
Canadienne	0,0000	0,0000	0,2105	0,4474	0,2895	0,0526
Holstein	0,0135	0,0000	0,4324	0,4189	0,1351	0,0000
Suisse-Brune	0,0000	0,0217	0,3478	0,5652	0,0652	0,0000

HEL 1

	Allèles					
	3	4	5	6	7	8
Canadienne	0,3333	0,0833	0,0000	0,0278	0,5278	0,0278
Holstein	0,3939	0,0909	0,0303	0,0303	0,4394	0,0152
Suisse-Brune	0,2391	0,3696	0,0000	0,0000	0,3044	0,0870

INRA 32

	Allèles						
	-1	1	2	3	4	5	6
Canadienne	0,0263	0,1316	0,1053	0,3684	0,2105	0,1316	0,0263
Holstein	0,0000	0,0000	0,2973	0,1216	0,4460	0,0405	0,0946
Suisse-Brune	0,0000	0,0652	0,0870	0,0652	0,4130	0,3478	0,0217

INRA 72

	Allèles			
	3	5	8	10
Canadienne	0,0000	0,0833	0,5278	0,3889
Holstein	0,0541	0,1216	0,2838	0,5405
Suisse-Brune	0,0000	0,1957	0,6739	0,1304

HEL 5

	Allèles							
	1	2	3	4	6	8	9	10
Canadienne	0,0263	0,0526	0,1842	0,2632	0,0000	0,0263	0,3421	0,1053
Holstein	0,0000	0,0135	0,0541	0,5676	0,0135	0,0000	0,3514	0,0000
Suisse-Brune	0,0000	0,2391	0,1522	0,0217	0,0000	0,0000	0,5435	0,0435

MM 8							
	1	2	3	Allèles 4	5	6	7
Canadienne	0,0000	0,0526	0,2105	0,1053	0,5000	0,0000	0,1316
Holstein	0,0429	0,1429	0,1857	0,0143	0,4286	0,0143	0,1714
Suisse-Brune	0,0000	0,1304	0,3696	0,0870	0,3261	0,0000	0,0870

INRA 16										
	-2	-1	1	2	Allèles 3	4	5	6	7	8
Canadienne	0,0000	0,0000	0,0000	0,0263	0,0263	0,1579	0,0000	0,0000	0,0263	0,7632
Holstein	0,0143	0,1714	0,0286	0,0857	0,0000	0,2857	0,0143	0,0429	0,0000	0,3571
Suisse-Brune	0,0000	0,0227	0,0000	0,0000	0,0227	0,3182	0,0000	0,0455	0,0682	0,5227

Annexe 2. Distances génétiques de Cavalli Sforza

	Canadienne	Holstein	Suisse-Brune
Canadienne	0,0000	0,0489	0,0297
Holstein	0,0489	0,0000	0,0597
Suisse-Brune	0,0297	0,0597	0,0000

Annexe 3. Distances génétiques de Nei

	Canadienne	Holstein	Suisse-Brune
Canadienne	0,0000	0,1976	0,1303
Holstein	0,1976	0,0000	0,3117
Suisse-Brune	0,1303	0,3117	0,0000

Annexe 4. Identification des animaux utilisés

Races					
Canadienne		Suisse-Brune		Holstein	
No. d'identification utilisés lors de l'étude	No. d'identification du PATLQ	No. d'identification utilisés lors de l'étude	No. d'identification du PATLQ	No. d'identification utilisés lors de l'étude	No. d'identification du PATLQ
C1	CN63615	SB1	BS32974	32	5192886
C2	CN65006	SB2	BS30617	98	5287164
C3	CN65151	SB3		169	5546580
C4	CN64544	SB4	BS32970	958	
C5	CN64628	SB5	BS30938	53	5251244
C6	CN65178	SB6	BS33401	24	5192851
C7	CN65181	SB7	BS30396	37	5200154
C8	CN65296	SB8		F9118	5416188
C9	CN65332	SB9	BS31633	F6013	4464060
C10	CN61925	SB10	BS34851	F7081	5447647
C11	CN64848	SB11	BS37006	F8207	4112653
C12	CN62884	SB12	BS30202	G339	
C13	CN65227	SB13	BS35447	G363	
C14	CN65133	SB14	BS33430	G2114	5241964
C15	CN63831	SB15	BS32973	S2	4412046
C16	CN64483	SB16	BS34574	S10	5051788
C17	CN63374	SB17	BS30503	S15	5594476
C18	CN63849	SB18	BS32406	S19	5485756
C19	CN64454	SB19	BS30614	S102	5341367
		SB20	BS30852	D28	4811056
		SB21		M734	4670321
		SB22	BS31461	M803	5349607
		SB23		M993	5861643
		Pistil	BS32512	M1041	5956375
				M1103	5513073
				M1189	5832405
				M1195	6386303
				H4	5299102
				H7	5832168
				H11	5843684
				H13	5424951
				H14	5757743
				H16	5578901
				H21	5928140
				H22	5177916
				H24	5895557